# EU liability rules for the age of artificial intelligence

Buiten , Miriam C.; DE STREEL, Alexandre; Peitz, Martin

*Publication date:*
2021

*Document Version*
Publisher's PDF, also known as Version of record

cerre Centre on Regulation in Europe

REPORT

**March 2021**

Miriam Buiten
Alexandre de Streel
Martin Peitz

# EU LIABILITY RULES FOR THE AGE OF ARTIFICIAL INTELLIGENCE

*As provided for in CERRE's bylaws and procedural rules from its "Transparency & Independence Policy", all CERRE research projects and reports are completed in accordance with the strictest academic independence.*

*The project, within the framework of which this report has been prepared, was supported by the following members of CERRE: Vodafone Group and Google. However, they bear no responsibility for the contents of this report. The views expressed in it are attributable only to the authors in a personal capacity and not to any institution with which they are associated. In addition, these views do not necessarily correspond either to those of CERRE, or of any sponsor or of any other member of CERRE.*

*ACKNOWLEDGEMENTS*

*The authors would like to thank the CERRE Secretariat, as well as the external representatives who participated in discussions, including representatives of BEUC, DG JUST, DG CNECT and DG GROW.*



*info@cerre.eu*
*www.cerre.eu*

# Table of contents

## About CERRE

Providing top quality studies and dissemination activities, the Centre on Regulation in Europe (CERRE) promotes robust and consistent regulation in Europe's network and digital industries. CERRE's members are regulatory authorities and operators in those industries as well as universities.

CERRE's added value is based on:

- its original, multidisciplinary and cross-sector approach;
- The academic qualifications and policy experience of its team and associated staff members;
- its scientific independence and impartiality;
- the direct relevance and timeliness of its contributions to the policy and regulatory development process applicable to network industries and the markets for their services.

CERRE's activities include contributions to the development of norms, standards and policy recommendations related to the regulation of service providers, the specification of market rules and the improvement of infrastructure management in a rapidly changing social, political, economic and technological environment. The work of CERRE also aims to refine the respective roles of market operators, governments and regulatory bodies, as well as aiming to improve the expertise of the latter, given that - in many Member States - the regulators are relatively new to the role.

## About the authors

**Miriam Buiten** is a CERRE Research Fellow and Assistant Professor of Law and Economics at the University of St.Gallen, Switzerland. She leads a research team on "Platform Governance", funded by the University of St.Gallen Basic Research Fund. Her research focuses on the legal issues surrounding new technologies and artificial intelligence and the role of competition law in regulating the digital economy. Previously, Miriam was a Junior Professor of Law and Economics at the University of Mannheim. She has been involved in several policy studies for the European Commission and the Dutch government on topics such as the role of online intermediaries in the ecommerce sector and mechanisms to reduce regulatory burdens.

**Alexandre de Streel** is Academic Co-Director at CERRE and a professor of European law at the University of Namur and the Research Centre for Information, Law and Society (CRIDS/NADI). He is a Hauser Global Fellow at New York University (NYU) Law School and visiting professor at the European University Institute, SciencesPo Paris and Barcelona Graduate School of Economics, and also assessor at the Belgian Competition Authority. His main areas of research are regulation and competition policy in the digital economy as well as the legal issues raised by the developments of artificial intelligence. Recently, he advised the European Commission and the European Parliament on the regulation of online platforms. Previously, Alexandre worked for the Belgian Deputy Prime Minister, the Belgian Permanent Representation to the European Union and the European Commission (DG CNECT). He holds a Ph.D. in Law from the European University Institute and a Master's Degree in Economics from the University of Louvain.

**Martin Peitz** is a CERRE Research Fellow and Professor of Economics at the University of Mannheim. He is also a Director of the Mannheim Centre for Competition and Innovation. His policy research focuses on digital markets, regulation, and competition economics. Martin holds a PhD in Economics from the University of Bonn.

# EXECUTIVE SUMMARY

# Executive summary

Two questions about the liability of Artificial Intelligence (AI) deserve attention from policymakers: 1) Do existing civil liability rules adequately cover risks arising in the context of AI systems? 2) How would modified liability rules for producers, owners, and users of AI play out? The report addresses the two questions for EU non-contractual liability rules while acknowledging the interaction of such rules with other regulatory instruments. The normative framework of the report is provided via an economic analysis of the law, considering how liability rules affect the incentives of producers, users, and other parties that may be harmed.

## 1. Opportunities and risks associated with AI and challenges for liability rules

Advanced AI systems, equipped with learning abilities, can improve human decision-making by approaching problems in fundamentally different ways. In doing so, these **AI systems promise to improve societal well-being and save time and resources in numerous fields**, including healthcare, transport, and consumer products.

As a result of the different approaches to problem-solving, a possible downside of AI is that in the event that errors occur in AI systems, these errors may be **less predictable** to humans. As AI systems advance, they will increasingly be used to support and – possibly, in some cases - replace human decision-making. It may then be unclear under which circumstances a human supervisor should intervene and override the decision taken by the AI system.

When AI systems possess the characteristics of unpredictability and autonomy, they present challenges for the existing liability framework. Liability rules have to address **who bears responsibility for accidents in which AI systems are involved**. The arrival of autonomous or semi-autonomous **AI systems tends to shift control over these systems away from owners and users ("operators") and towards producers**. Given that producers of autonomous AI systems can exercise more control over the performance of these systems than producers of mechanical products, it may be justified to shift more responsibility towards producers of AI systems. While AI may lead to many benefits, it seems unlikely that Al producers will be able to completely prevent AI-inflicted injuries.

Whereas producers control the product's safety features and provide the interfaces between the product and its operator, **the operator exercises control over the use of the system**. The operator decides in which circumstances the system is used and is in a position to oversee it in real-world situations. It is therefore appropriate to attribute some liability to operators who choose to delegate decisions to AI systems. However, it may be **difficult to attribute fault** to operators when they could not have reasonably anticipated the actions of the AI system. The fault would need to be established in terms of a failure to maintain or oversee the AI system. As AI systems gain autonomy, the scope of a duty to supervise may not be clear and injured parties may be unable to prove the fault of the operator and fail to obtain compensation from them under fault-based liability. Injured parties may also **have difficulty proving causality** between the AI systems' actions and the harm.

## 2. Guiding principles for liability rules

Following an economic approach, the report identifies the following guiding principles for liability rules for AI:

- **EU non-contractual liability rules should not be thought of in isolation but as part of a broader set of rules** as they jointly shape the incentives of all parties; in particular, liability rules need to be coherent with EU exante rules on safety and surveillance, the national non-contractual and contractual liability rules and rules on insurance; they also need to provide for the optimal degree of harmonisation in the EU and, thereby, respect the principle of subsidiarity.

- As always with regulatory design, **rules reflect trade-offs which should be well-identified**; liability rules address possible trade-offs between the interests of the producers (and their innovation) and the interests of the users (and their protection); the level of harmonisation of liability rules and the scope of such harmonised rules present trade-offs between ensuring legal certainty for injured parties and operators with a uniform framework, while preserving the internal coherence of Member States' national liability rules allowing for learning effects to be delegated to the Member States;
- **Safety is a shared responsibility** and tort liability should provide incentives to all stakeholders (producers, operators and users) to take an efficient level of care in designing, testing and employing AI-based solutions, recognising that care by each party may be essential to avoid a failure (complementary efforts); liability rules should also place liability on the least cost avoider, i.e. the party that can reduce harm at the lowest cost, acknowledging that a party's incentives are determined by its private costs and benefits of reducing harm, which may depend on the decisions taken by others;
- Liability rules should be **based on risks of harm**, which may differ depending on the application and the context in which AI systems are used;
- Liability rules should ensure an **efficient disclosure of information** in situation of information asymmetries between stakeholders;
- Liability rules should balance **proactive policymaking**, anticipating technological changes, with **reactive policymaking**, adapting the rules only after having gained some experience from deploying the technologies;
- Liability rules should be **principles-based and flexible** while allowing for sufficient legal certainty and predictability for all stakeholders;
- Based on consequentialist ethics, liability rules should be **technologically neutral**, providing the same level of protection of users of a product or services powered by AI as users of the same type of product or service which is not powered by AI (however, society may want to follow different ethics and, thus, depart from technological neutrality).

## 3. Adapting the EU liability rules to AI challenges

**The report identifies three dimensions relevant to reviewing the EU liability framework for AI systems: (1) who should be liable; (2) the scope of new rules; and (3) the level of harmonisation.**

On the question of who should be liable, the report considers the liability of producers and of operators. Given the upcoming **review of the EU Product Liability Directive**, the report considers what challenges posed by AI could be addressed within the Product Liability Directive. **The report makes four main points in this regard**.

- The report recommends **clarifying that software is included** in the definition of a product; in the age of digitalisation, differentiations between tangible and intangible objects of use are more difficult to justify;
- Concerning the safety expectations consumers are entitled to have of a product, a reform should **consider the dynamic nature of software products, IoT devices and AI systems**;
- The notion of the defect should be clarified, recognising that autonomous AI systems make it difficult to draw the line between acceptable autonomous behaviour and unacceptable errors. In the context of autonomous AI systems, the notion of **defect may need to be defined in terms of overall failure rate rather than individual error**;

- The **standard of proof** for proving a defect and the causality between the defect and the harm **may need to be lowered** to facilitate injured parties obtaining compensation for harm;

Next to manufacturers, the report identifies several **reasons to keep operators of AI systems accountable**. Firstly, liability for operators encourages them to take precautions in supervising AI systems that are not fully autonomous. Secondly, for highly autonomous AI systems, liability provides an incentive for operators to keep the system updated and ensure that it is used properly. Thirdly, operators benefit from employing AI, making it appropriate for its costs to be internalised.

In terms of liability standards that should apply to operators, the report acknowledges that introducing strict liability for AI would constitute a sharp departure from the standard liability regime currently in place in several Member States. Given that the risks of employing AI systems depend on the type of device and the context in which it is used, it is recommended to consider **strict liability only in certain sector-specific or application-specific contexts. Lowering the standard of proof is an alternative means** to facilitate injured parties' access to compensation, which is already happening at Member State level in other contexts.

The report considers to what extent the characteristics of AI justify more EU harmonisation of liability rules, beyond the context of producer liability. Considering that the diversity of rules between the Member States allows for experimentation and learning and that further harmonisation would interfere with the internal coherence of Member States' liability regimes, **harmonisation may be limited to the sectors where there is a set of EU safety rules that liability rules may usefully complement**. The report concludes that the need for an across-the-board harmonised regime for AI at the European level can be questionable. Instead, the report proposes harmonising EU liability rules with existing sector-specific rules, for three main reasons: 1) introducing a harmonised operator liability regime for AI systems, or high-risk AI systems, may lead to delimitation difficulties. It may be difficult to find a clear-cut, yet general criterion for distinguishing between "ordinary" and "autonomous" systems. 2) listing "high risk" AI applications may presuppose that AI applications create similar risks regardless of the context in which they are applied. Such a regime would therefore need to define not only the high-risk technologies but also the applications or contexts that it covers. 3) existing sector-specific regulation already reflects the need to differentiate regulation according to the context in which technology is applied. Overall, it appears that many AI applications would fall in the non-high-risk category, and those that are high-risk will predominantly be covered by sector-specific regulation. The added value of a horizontal liability regime for high-risk AI, as compared to specifying liability rules in sector-specific regulation, may therefore be limited if sector-specific regulation is adequate.

# 01

# INTRODUCTION

# 1. Introduction

The rapid advancements in Artificial Intelligence (AI) offer myriad opportunities for society. Machine Learning (ML) algorithms are all around us, driving numerous tools and applications used in everyday life.[1] Robotic household appliances and personal digital assistants save consumer's time and effort, advanced search algorithms serve them personalised products, services, and content. By supporting and replacing human decision-making, AI also promises to reduce injury and harm from accidents in a wide range of contexts. AI tools are already improving diagnostics in healthcare,[2] and (semi-) autonomous cars promise to increase road safety.[3]

These AI systems[4] can improve human decision-making by approaching problems in fundamentally different ways. ML algorithms are considerably better at recognising patterns than humans. Yet, as a result of this different approach, AI errors may be less predictable to humans.[5] Where AI supports decision-making, it may also be unclear in what situations a human supervisor should intervene and override the decision. For instance, when is a physician allowed to rely on a cancer detection tool? When should the driver of an autonomous car be paying attention and take over the driving?

In such cases, AI systems present legal challenges for the existing liability framework. In its White Paper on Artificial Intelligence and its Report on the safety and liability implications of AI,[6] the **European Commission suggests a risk-based approach to regulating AI**. The Commission envisages new regulatory provisions and preliminary assessments to ensure that "high risk" AI systems meet the requirements of security, fairness, and privacy. The European Commission's intends to meet two objectives: On the one hand, to create an "ecosystem of excellence" along the entire value chain, starting with research and innovation, and to provide the right incentives to accelerate the implementation of AI systems, including by small and medium-sized enterprises (SMEs). On the other hand, an "ecosystem of trust" is to be established to ensure compliance with EU regulations and the protection of citizens/consumers.[7] On that basis, the Commission envisages several options for reform such as better enforcement (requiring more transparency), adaptation of the scope of EU legislation (in particular regarding software products), as well as clarification and change in the allocation of responsibilities between the different actors involved in the AI lifecycle. The EU Commission further suggests the adoption of additional rules and a strict liability regime for high-risk AI. The Commission is expected to announce its proposals for reform on 21st April 2021.[8]

In October 2020, the **European Parliament** approved three resolutions on AI covering ethics, civil liability, and intellectual property. The resolution with recommendations to the Commission regarding a civil liability regime[9] **strives for more harmonisation and, like the AI White Paper, follows a risk-based approach**. To address the challenges posed by emerging digital technologies, the resolution proposes to revise the Product Liability Directive, while complementing the existing fault-based tort law of the Member States in specific cases, particularly where third parties causing harm are untraceable or impecunious.

---

[1] See e.g. Domingos (2015).

[2] For example, tools for cancer detection, see <https://eujournal.org/index.php/esj/article/view/8693>, <https://www.nature.com/articles/d41586-020-03157-9>, <https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30003-0/fulltext>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6646649/>.

[3] See e.g. Teslka Vehicle Safety Report, <https://www.tesla.com/VehicleSafetyReport>, Waymo's Safety Report, <https://waymo.com/safety/>.

[4] The report will use the terms "AI systems", "AI applications", "AI devices" and "AI tools" interchangeably.

[5] Yoshikawa (2019), p. 1163.

[6] Communication White Paper of 19 February 2020 on Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65 (henceforth, "**AI White Paper**") and Commission Report of 19 February 2020 on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, COM(2020) 64 (henceforth, "**AI Commission Report**").

[7] AI White Paper.

[8] See https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements.

[9] European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)) (henceforth, "**EP Resolution**").

For autonomous AI systems that entail inherently high risk, the European Parliament resolution proposes a strict liability regime on the EU level, coupled with mandatory insurance.[10] For other AI systems, the European Parliament resolution proposes a fault-based liability rule with a presumption of fault on the part of the operator.[11]

**To inform these policy initiatives, our report aims to identify the gaps in existing liability rules in cases involving AI systems, in particular ML technologies, to assess if and how these liability rules may need to be reviewed.** The economic analysis of law provides the normative framework of the report and considers how liability rules affect the incentives of producers, users, and other parties that may be harmed.

The **scope of the report is limited to EU non-contractual liability rules for AI.** We recognise that tortious liability is part of a broader regulatory framework that includes ethical guidelines, safety regulations, and rules on contractual liability. We acknowledge that the European Commission considers these and other angles next to liability rules to address any concerns related to the use of AI. While Member States may have different national regulatory solutions in place, we will refer to them only where relevant as major differences exist. The report will focus on AI harm that falls within the ambit of liability rules.[12] The report excludes the discussion on AI as a legal person, following the recommendations of the Expert Group report[13] and the European Parliament resolution.[14]

This report proceeds as follows: Section 2 lays out the current legal framework for AI applications, illustrating that non-contractual liability rules are part of a broader regulatory framework.

In Section 3, the report considers what **new risks AI introduces** as compared to non-AI tools. Building on the Expert Group report, the report identifies the characteristics unique to AI technologies.15 By illustration of concrete examples, the report considers what **challenges these risks present to the liability regime**. We distinguish three questions:

- How is responsibility divided among actors involved? Do the characteristics of AI provide reasons to shift liability to different parties because AI may shift control to a different party as compared to non-AI tools?
- What is the standard for liability? Do the characteristics of AI provide reasons to move from a fault-based standard to a strict liability regime, for instance, because we may find that AI users should compensate harm even if we are unable to establish fault on their part when employing AI?
- What do injured parties need to prove? Do the characteristics of AI provide reasons to change procedural rules? Primarily, we discuss if we need to reconsider rules on burden of proof in the context of AI because it arguably becomes more difficult to identify the cause of harm.

[10] EP Resolution, Art. 4.
[11] EP Resolution, p. 38.
[12] We acknowledge that there are broader concerns surrounding AI and its impact on society, including the role of algorithms in misinformation and polarisation, the potential addictive aspects of attention platforms, the impact of algorithm-based online platforms on, for instance, the labour market. There may be a need for ethical rules or certification requirements to address these broader societal concerns: see Belfield et al. (2020). See also Alter (2017); Williams (2018).
[13] Expert Group on Liability and New Technologies New Technologies Formation, Liability For Artificial Intelligence And Other Emerging Digital Technologies (2019) (henceforth, "**Expert Group Report**")
[14] EP Resolution, p. 6, number 7.
[15] Expert Group on Liability and New Technologies Formation, Liability For Artificial Intelligence And Other Emerging Digital Technologies 2019.

Having identified the points of legal tension, Section 4 analyses the incentive effects of liability rules for the development and employment of AI. Based on the economic approach, the report identifies a set of **guiding principles for liability rules for AI**.

In Section 5, the report analyses **what would be the best approach in addressing any gaps in the liability rules**. The report considers three dimensions: (i) Who should be liable? The report discusses what challenges of AI could be resolved by revising the product liability rules, and what problems need to be addressed in the general rules on fault-based liability; (ii) What should be the scope of the new rules? The report considers whether new liability rules should apply to all AI, to specific types of AI, or follow sector-specific rules; (iii) What level of harmonisation is required? So far, the Product Liability Directive is an exception of liability rules at EU level. The report discusses whether AI provides reasons to change this approach and harmonise a larger part of liability rules at EU level.

# 02

# EXISTING LEGAL FRAMEWORK RELEVANT TO AI

# 2. Existing legal framework relevant to AI

## 2.1. Different liability standards

As explained by the Commission, 'civil liability rules play a double role in our society: on the one hand, they ensure that victims of a damage caused by others get compensation and, on the other hand, they provide economic incentives for the liable party to avoid causing such damage'. The Commission also notes that: 'liability rules always have to strike a balance between protecting citizens from harm while enabling businesses to innovate'.[16]

Different standards of liability can be seen as a continuum, ranging from less to more favourable to victims.

### *a. Usual standard: Fault-based liability without presumption*

The fault-based liability regime is the standard liability applicable in Member States. Under such a regime, claimants need to prove three cumulative conditions to get damages:

- *Fault*: Fault results from someone failing to act as could be expected from a reasonable person. Fault consists of either a violation of the law or something that is against the normal cautiousness that can be expected. The diligence required is higher for professionals.
- *Damage:* To recover damages, the plaintiff must prove that the culpable conduct by the defendant resulted in damage, and, in some legal systems, that the legally protected interests of the plaintiff were violated. Damages may cover material and immaterial damages.
- *Causality:* For a causal link to be established, it often suffices that one element contributed to the damages. In the case of complementary products/services, it suffices that one contributed to the damages to be responsible for the entire damages.

As the proof of each of those conditions may be difficult, some other – more victim-friendly – liability regimes have been established in Member States for specific situations where the legislator (or the case-law) has estimated that the victims need to be better protected.

### *b. Fault-based liability with presumptions*

A first victim-friendly regime consists of **starting from the standard fault-based regime but changing the standard of proof** or some conditions. A rebuttable presumption for the fault requirement and/or for the causal link can facilitate victims in obtaining compensation, and/or can help reduce information asymmetry between the victim and the wrongdoer.

As explained by the Commission,[17] a **presumption regime may be linked to a diverse set of factual situations generating different types of risks and damages**, such as (i) the responsibility of the owner/possessor of the building in case of damages causes by his/her building (unless he/she proves that he/she observed appropriate care to avoid the damage); (ii) the responsibility of a person carrying out a dangerous activity (unless he/she proves that all appropriate measures to avoid the damage have been taken); (iii) the responsibility of the employer/the principal for the actions executed on his behalf or interest by his employees/agents (unless he proves that he used appropriate care in the selection and the management of the agent/employee) or (iv) the responsibility of parents/tutors/guardians/teachers for damages caused by a minor, pupil,

[16] AI Commission Report, p. 12.

[17] Commission Staff Evaluation of 7 May 2018 of Council Directive 85/374 of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products, SWD(2018) 157 (henceforth, "**PLD Commission Evaluation**"), p. 7.

student/apprentice, or mentally impaired person (unless they can prove that they were not able to prevent the damages from happening).

### *c. Strict liability*

A second more victim-friendly regime consists of facilitating the proof of the victim by changing the conditions which needed to be proven to get damages. Under a strict liability regime, **victims need to prove**:

- the *default* or the risks taken by the wrongdoers which is easier to prove than the fault - or the negligence - of the wrongdoer;
- the *damages*; which types are often limited or capped;
- the *causality link* between the default/risks taken and the damages; note that the need to prove the causality link may be an important obstacle to the victims.[18]

As explained by the Commission, a special standard may be justified because: [19]

- the risk of damage is linked to the unpredictability of behaviour of specific risk groups, like animals or certain persons: in these cases, liability may be attributed to the persons that are considered responsible to supervise the animal or the person, because it is them who should normally be in the condition to adopt measures to prevent or reduce the risk of damages.
- the risk of damages is linked to dangerous activities: some jurisdictions may attribute liability to the person that carries out the activity (e.g. the operator of a nuclear power plant or an aircraft or the driver of a car) or are ultimately responsible for the dangerous activity to happen (e.g. the owner of a vehicle). The rationale typically is that this person has created a risk, which materialises in damage and at the same time also derives an economic benefit from this activity.

The Commission also explains that those **strict liability regimes may apply to a diverse set of factual situations generating different types of risks and damages**, such as (i) the liability of the owners of animals for the damages caused by the animals under their custody, (ii) the strict liability of the person responsible for carrying out an unspecified or specified dangerous activity (for example the operation of nuclear power plants, aircraft, or motor vehicles or (iii) other cases linked to a legal or factual relationship between two persons or a person and an object, such as when the damages are caused by someone executing a task in the interest of someone else (employee/employer) or by an object that is under his/her custody.

**Some forms of strict liability may go even a step further by linking liability simply to the materialisation of risk** and/or making the discharge of liability either impossible or possible only under the proof that the damaging event was caused by an exceptional/unforeseen circumstance that could not be avoided. In effect, those stricter regimes establish presumptions of a causality link to facilitate the compensation of the victim of damages in situations where the legislator considers it too burdensome or unbalanced to require the victim to prove such causality link.

As the strict liability regime tilts the balance in favour of the victim at the expense of the person responsible, they are in general **accompanied with limiting principles, esp. regarding the type of damages which can be compensated or the maximum of damages which can be granted**. Thus, relying on a strict liability regime (when it exists), the victim may be compensated more easily than under the fault-based liability regime but a more limited manner. If the victim

[18] Commission Report of 7 May 2018 on the Application of the Council Directive on the approximation of the laws, regulations, and administrative provisions of the Member States concerning liability for defective products (85/374/EEC), COM(2018) 246 (henceforth, "**PLD Commission Report**"), pp. 5-6.
[19] PLD Commission Evaluation, p.8.

seeks compensation for more damages than the ones covered by strict liability, the victim needs to also launch a complementary action against the person responsible under fault-based liability. It is also interesting to note that strict liability regimes may also be coupled with mandatory insurance requirements.

#### *d. Several wrongdoers, joint and several liabilities*

Where more than one party is liable for compensation of the same damage, **tortfeasors are in general jointly and severally liable**. Redress claims between tortfeasors are usually possible for identified shares of the victim of damages in situations where the legislator considers it too burdensome or unbalanced to require the victim to prove such causality link.[20] Generally, groups will re-allocate the costs of liability by contractual agreement. Supply agreements among the manufacturers, i.e. end-producers and component suppliers, routinely include clauses that allocate the costs caused by defective components.[21]

## 2.2. The overall regulatory framework

The rules for **non-contractual liability are part of a broader legal framework, which includes other types of liability (notably contractual liability) as well as ex ante rules and guidelines (notably safety rules)**. Moreover, additional sector-specific liability and safety rules apply in high-risk sectors. This legal framework applies against the background of general ethical principles, such as those formulated in the High-Level Expert Group report, the OECD ethical principles and the G20 Ministerial Statement.

Overall, liability rules, which are ex post, are complementary with product safety rules, which are ex ante.[22] While product safety frameworks ensure that products entering the market are safe and that continued compliance with safety-requirements is guaranteed during their entire life-cycle, product liability takes effect retrospectively after the damage has occurred answering the question of who should be held responsible for a product that has caused damage.

**Table 1: Overall regulation framework**

| | **Ex ante safety rules + Surveillance** | **Ex post liability rules + Insurance** |
|---|---|---|
| **Horizontal** | GPSD + MSR | PLD |
| **Sector-specific (high risks)**<br>*e.g. Health care*<br>*Automotive*<br>*Machinery* | MDR and IVRD<br>GVSR + AMSVR<br>MD | |

[20] Expert Group Report, p, 8.
[21] See also Wagner (2019a), p. 32 who notes: "The obvious tool for re-allocation of the costs of liability within one of the groups is a contractual agreement. Already today, standard supply agreements among the members of the manufacturer group, i.e. end-producers and component suppliers of different layers routinely include clauses that provide for the allocation of the costs of product recalls and other costs caused by defective components."
[22] AI Commission Report, p. 12.

### 2.2.1. Ex ante safety rules

#### 2.2.1.1. Horizontal rules

The mass-manufacturing of products for profit carries an inherent risk of damaging the consumer. Nevertheless, the commercialisation of goods remains worthwhile as it brings along substantial overall social benefits. However, certain risks are deemed unacceptable and hence are addressed ex ante through **product safety legislation**. These rules set the standard for the essential safety requirements having to be met by a product when entering the European market.[23]

Within the EU, the product safety rules are divided into two levels of legislation. Specific rules regulate certain sectors or products, and in absence of such specific requirements, the general rules set out in the **General Product Safety Directive (henceforth, "GPSD")**[24] apply. The directive ensures that only **safe consumer products** (i.e. that *do not present any risk or only the minimum risks* under normal conditions of use taking into account, *inter alia,* its characteristics and effects of the product on other products[25]) are placed on the market by manufacturers.[26] In particular, the GPSD imposes an obligation on producers as well as distributors to not only provide safe products to consumers but further to take all possible steps to identify any hazards of their products and to inform consumers of the existence of such risks. Furthermore, producers and distributors are obliged to monitor the market to take possible necessary measures, to deal with unidentified risks, or to withdraw dangerous products.[27]

The legislative framework is limited to defining only the essential safety requirements. The specific technical specifications, however, are elaborated by standards developed by Standards Setting Organisations (SSO) at the international level (e.g. ISO), European level (CEN-CENELEC-ETSI), or national level (e.g. UNI, DIN).[28] Even though these standards are in general not legally binding, they nevertheless remain important as they can act as an indicator on whether a product is to be deemed safe or not. In that sense, a product is presumed to be safe if it is manufactured under European technical standards.[29]

The European Commission is preparing a review of the GPSD that may result in a proposal for a revision by the second quarter of 2021.[30] Part of the initiative is to tackle the product safety challenges arising from the development of new technologies.

The **Market Surveillance Regulation** sets out requirements for accreditation/market surveillance and confers powers to national authorities.[31] EU market surveillance legislation provides (i) clear and uniform rules applying to non-food products and economic operators, (ii) requirements in terms of infrastructure, organisation, legal powers to ensure that market surveillance can cope with enforcing EU legislation, (iii) streamlined market surveillance procedures for controlling products within the EU and at its borders and (iv) tools to coordinate activities carried out by national

[23] Bertolini (2020), p. 50.
[24] Directive 2001/95 of the European Parliament and of the Council of 3 December 2001 on general product safety, OJ L 11 of 15.01.2002.
[25] Article 2 GPSD.
[26] Article 3 GPSD.
[27] Article 3 GPSD.
[28] For an overview, see https://ec.europa.eu/growth/single-market/european-standards/harmonised-standards_en. See further Timan et al. (2019); AI Commission Report, pp. 3 ff.; and, more in detail, Commission Notice of 26 July 2016, The "Blue Guide" on the implementation of EU product rules 2016, C 272/1.
[29] Articles 3(2), 2(2) and 3(3) GPSD.
[30] See https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12466-Review-of-the-general-product-safety-directive.
[31] Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011, OJ L 169 on 25.6.2019. This Regulation will be applicable starting July 2021 and replaces Regulation 765/2008 of the European Parliament and of the Council of 9 July 2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products and repealing Regulation (EEC) No 339/93, OJ L 218 on 13.8.2008.

surveillance bodies across the EU such as discussion forums, IT databases, and common market surveillance campaigns.[32]

#### *2.2.1.2. Sector-specific regulation for high-risk sectors*

For high-risk sectors, the EU product safety framework complements the horizontal rules with sector-specific rules. Concerning AI applications, three sector-specific regulation frameworks require special attention.

AI is widely employed in **healthcare**,[33] for instance in the form of predictive algorithms precision medicine to support in diagnosing, selecting drugs or prioritising patients.[34] Medical devices within the EU are about to be regulated by new regulations which came into force in May 2017 and become valid with graduated transition periods of 6 months to 5 years: the Medical Devices Regulation (henceforth, "MDR")[35] and the In Vitro Diagnostic Medical Devices Regulation (henceforth, "IVDR").[36]

In the domain of **transportation**, in particular autonomous cars, the following specific regulation relevant to autonomous cars applies: the General Vehicles Safety Regulation (henceforth, "GVSR")[37], the Approval and Market Surveillance of Vehicles Regulation (henceforth, "AMSVR")[38] and the Motor vehicles Insurance Directive (MID).[39] Those rules are explained below in Section 3.3.1.

For AI integrated into **machinery**,[40] the Machinery Directive[41] applies and a CE Declaration is required. The directive promotes the free movement of machinery within the single market and guarantees a high level of protection for EU workers and citizens. As it is a 'new approach' directive, it promotes harmonisation through a combination of mandatory health and safety requirements and voluntary harmonised standards.

### *2.2.2. Ex post liability rules*

#### *2.2.2.1. EU law: Product Liability Directive*

The **EU Product Liability Directive** (henceforth, "PLD"), adopted in 1985, established a strict liability regime where producers are liable for their defective products regardless of whether the defect is their fault.[42] The PLD empowers consumers to obtain compensation for damage caused by defective products.

---

[32] <https://ec.europa.eu/growth/single-market/goods/building-blocks/market-surveillance_en#market_surveillance_legislation>.

[33] Tata Consultancy Services (2017). Getting smarter by the sector: How 13 Global Industries Use Artificial Intelligence, <https://sites.tcs.com/artificial-intelligence/>.

[34] Price (2017).

[35] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC, OJ L 117 on 5.5.2017.

[36] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU, OJ L 117 on 5.5.2017.

[37] Article 3(22) GVSR: "a motor vehicle that has been designed and constructed to move autonomously without any driver supervision".

[38] Regulation 2018/858 of the European Parliament and of the Council of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, OJ L 151 of 14.06.2018. This Regulation took force on 1 September 2020 and repeals Directive 2007/46 of the European Parliament and of the Council of 5 September 2007 establishing a framework for the approval of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles (Framework Directive).

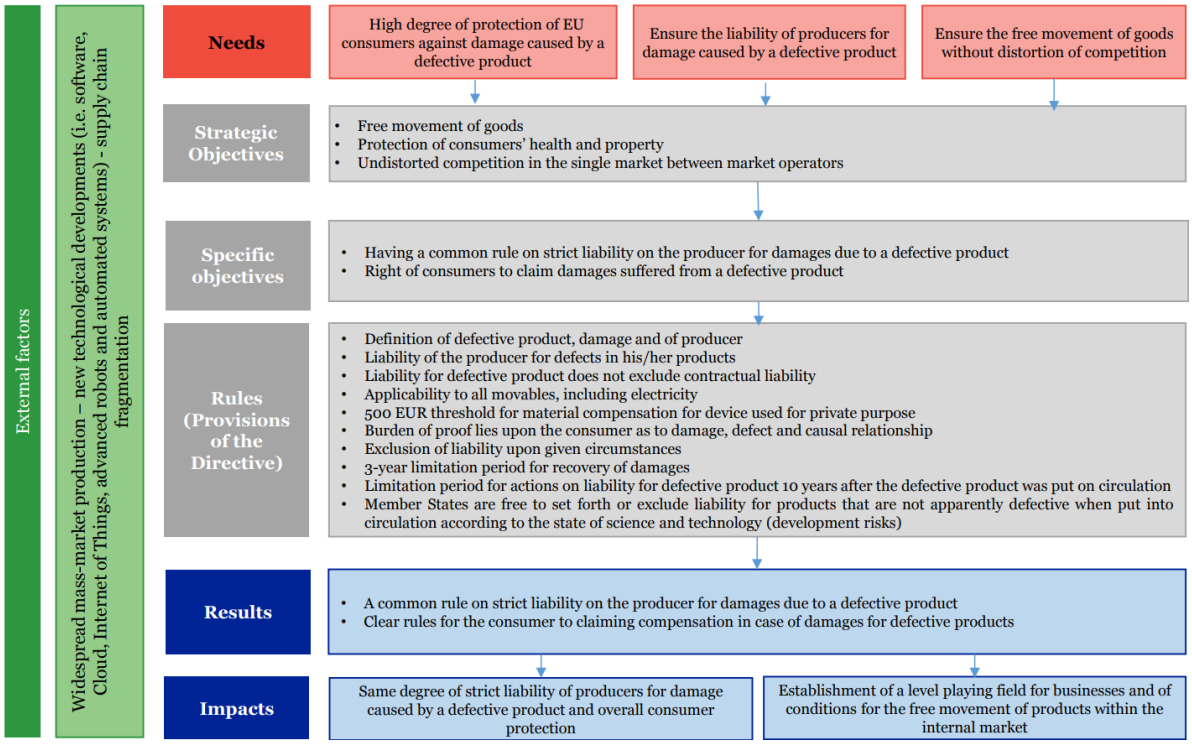
[39] Directive 2009/103 of the European Parliament and of the Council of 16 September 2009 relating to insurance against civil liability in respect of the use of motor vehicles, and the enforcement of the obligation to insure against such liability.

[40] Machinery defined in Article 1 and 2(a) Machinery Directive.

[41] Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (recast), OJ L 157 on 9.6.2006

[42] Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products.

**Figure 1: Intervention logic of the PLD**




*Source: PLD Commission Services Evaluation, p. 7*

The PLD is a technology-neutral instrument that fully harmonises product liability rules throughout the EU. It applies to any **product** sold in the EEA with a three-year limit for the recovery of damages. The PLD applies to "movables" (Art. 2), which are interpreted as tangible goods.[43] The Court has indicated that the PLD applies to products used while providing any service.[44] Traditionally, imposing strict liability for product defects, not on services,[45] is justified by it being easier for a plaintiff to prove a defendant was negligent concerning their behaviour than to provide evidence about the defective nature of a product.[46] This means that a product can include digital content if it is embedded into a tangible good, such as is usually the case for IoT products. However, views on the legal classification of digital content vary. This means that it is unclear if the PLD applies to software that consumers purchase separately.[47]

The PLD assigns liability to the "**producer**" (Art. 1 PLD). It defines "producer" as the manufacturer of a finished product, the producer of any raw material or the manufacturer of a part, and any person who, by putting his name, trademark or other distinguishing feature on the product presents himself as to its producer. Any person who imports a product for sale, hire, leasing or any form of distribution in the course of his business is responsible in the same way as the producer (Art. 3(2) PLD).[48]

Producer liability under the PLD arises in case of a **defect**. The PLD defines a "defective" product as a product that does not provide the safety the consumer is entitled to expect, considering all

[43] See e.g. BEUC (2020), p. 12.
[44] Judgment of 10 May 2001, *Veedfald*, C-203/99, ECLI:EU:C:2001:258.
[45] Ebers (2020), p. 58. See also Judgment of 21 December 2011, *Dutreux*, C-495/10, ECLI:EU:C:2011:869; PLD Commission Evaluation, p. 7.
[46] See e.g. Carpenter & Collins (2012).
[47] BEUC (2020), p. 12 notes that "A distinction is sometimes made in the academic literature between "standardised software" and "bespoke software": standardised and mass-produced software is usually considered as "a product", while bespoke software is seen as an "individualised service", referring to Fairgrieve (2019); Howells et al. (2017); Wagner (2019a), p. 42. See further Section 5.2.2.
[48] BEUC (2020), pp. 18-19.

circumstances. This includes, for instance, the presentation of the product, the use to which it could reasonably be expected that the product would be put, the time when the product was put into circulation. The notion of defect focuses on consumers' safety expectations to physical harm, excluding possible privacy harm, cybersecurity flaws, or other risks that may arise for IoT products.[49] A defect shall be assessed considering "the time when the product was put into circulation" (Art. 6(1)(b) PLD). This concept has raised interpretative questions.[50]

The PLD defines **damage** as death, personal injury, or damage to the product or other property with a ceiling.[51] Thus, the PLD limits the type of damages which can be compensated as it does not cover immaterial damages. It also allows Member States to cap the number of damages.[52]

There is room for different national approaches, for example on systems to settle claims for damages, or on how to bring proof of damage. Member States may also introduce or maintain other national instruments for the liability of producers based on fault.[53]

#### *2.2.2.2. National laws: Fault-based liability and strict liability*[54]

The PLD empowers consumers to obtain compensation for damage caused by defective products. Each Member State has its tort law in place, as this is an area that is not harmonised on a horizontal level. These liability rules have in common that they are based on fault.[55] Member States differ in the requirements they impose for fault-based liability and the losses they recognise.

While it is not possible to engage in a comparative analysis of Member States' laws in this report, it can be said that **the rules on strict liability vary**. As an illustration, French law recognises strict liability in a wide variety of cases, whereas Germany only attributes strict liability in isolated cases.

Under French law, liability arises for any damage that arose as a result of the injurer's *faute* (Art. 1240-1241 French Civil Code). *Faute* requires behaviour that does not meet the standard of a just and cautious person. Damage may include pure economic loss. However, the damage must be direct, certain, and legitimate. **French law recognises a strict liability for the keeper (*guardian*) of a thing** (Art. 1242 French Civil Code), provided that the respective thing played an active role in creating the damage. The keeper is any person who possesses usage, control, and supervision of the good.[56] The keeper has the possibility of exonerating himself/herself by demonstrating contributory negligence of the victim, or a case of *force majeure*.[57] In the French literature, it has been argued that keepers of autonomous AI systems should not be held liable because they cannot control them.[58]

Under German law, negligence liability requires that one of the rights listed in § 823(1) BGB was infringed: life, health, property, freedom, personality, and commercial enterprise. Pure economic loss is generally not among damages compensated. Indirect losses can give rise to liability if the injurer breached a duty of care to prevent the damage ("Verkehrssicherungspflicht"). The injured party must prove causation, which requires foreseeability on the side of the injurer. **By exception, German law imposes strict liability for the keeper of motor vehicles** (§7(1) Strassenverkehrsgesetz) **and luxury animals** (§ 833(1) German Civil Code). The keeper is the person who benefits from the use of the good and who can control the object as a source of risk.

[49] BEUC 2020, p. 13.
[50] See e.g. Judgment of 9 February 2006, *Declan O'Byrne v. Sanofi Pasteur MSD*, C-127/04, ECLI:EU:C:2006:93.
[51] See BEUC 2020, p. 15.
[52] Art. 16 PLD.
[53] PLD Commission Report, p.4.
[54] Overview based on Janal (2020), p. 178 ff.
[55] Zweigert & Kötz (1996), p. 650.
[56] Janal (2020), p. 181 and the references therein.
[57] Janal (2020), p. 181 and the references therein.
[58] Janal (2020) refers to Mendoza-Caminade (2016), p. 447; Lagasse (2015).

For other "things", German law does not impose strict liability but recognises safety duties of the keeper under § 823(1) German Civil Code.

Whoever profits from the use of a hazardous object should bear the associated risk and is liable if they negligently violate safety duties ("Verkehrspflichten"). These safety duties include the obligation to monitor the object's status and activities and take reasonable preventive action to avert harm. Case law recognises several categories of hazardous objects, including buildings and premises, household chemicals, and washing machines.[59]

[59] See e.g. Förster (2017).

# 03

# CHALLENGES FOR LIABILITY RULES RAISED BY AI

# 3. Challenges for liability rules raised by AI

## 3.1. AI as a concept and risks associated with AI

### *3.1.1. Definition of AI*

**Artificial Intelligence is a general and broad term for various technologies, which have different features and are designed for different fields of application**. The concept of AI has been around since the late 1950s.[60] John McCarthy first coined the term artificial intelligence in 1956, defining AI as "the science and engineering of making intelligent machines, especially intelligent computer programs".[61] **Some contemporary definitions of AI focus on AI's ability to act autonomously,[62] or to evolve in an unforeseeable way.[63]** Definitions of AI focusing on the level of autonomy of the application may be interpreted differently depending on whom you ask, and when you ask it.[64] Other definitions consider AI as those applications that produce results that we perceive as achieving a level of human intelligence.[65] For instance, AI is an intelligent machine that "performs tasks that normally require human intelligence";[66] or AI are machines that "work to achieve goals".[67] Robots have been defined as systems "capable of perceiving the environment or context in which it is located, that can process the information to plan a certain action and execute it".[68]

AI is used as an umbrella term for various technologies that rely on algorithms. There appears to be consensus that not all algorithms constitute AI. Simple, rule-based algorithms are unambiguous specifications to solve a class of problems. More sophisticated, machine learning (ML) algorithms are programmed to find patterns in data through training, allowing them to identify the best possible model to explain the data. A subset of ML algorithms is neural networks, which allow various ML algorithms to collectively process complex data inputs. The possible applications are wide-ranging and include speech recognition, health diagnostic systems, self-driving cars, chatbots, and content recommendations.[69]

At EU level, the following definition of AI has been proposed at this stage. According to the European Commission, "artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications)."[70]

According to the AI High-Level Expert Group, "artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions." [71]

[60] Rachum-Twaig (2020) referring to: Minsky (1959); McCarthy (1959).
[61] McCarthy (2007), p. 1.
[62] Scherer (2016), p. 363.
[63] Gasser & Almeida (2017).
[64] Buiten (2019).
[65] Buiten (2019).
[66] Chung & Zink (2018).
[67] Scherer (2016).
[68] Calo (2015).
[69] See e.g. Stone et al. (2016).
[70] Communication from the Commission of 25 April 2018, Artificial Intelligence for Europe, COM (2018) 237, p. 1.
[71] High-Level Expert Group on Artificial Intelligence, A Definition of AI, 8 April 2019.

According to the European Parliament, "AI-system" means a system that is either software-based or embedded in hardware devices, and that displays behaviour simulating intelligence by, inter alia, collecting and processing data, analysing and interpreting its environment, and by taking action, with some degree of autonomy, to achieve specific goals and 'autonomous' means an AI-system that operates by interpreting certain input and by using a set of pre-determined instructions, without being limited to such instructions, despite the system's behaviour being constrained by, and targeted at, fulfilling the goal it was given and other relevant design choices made by its developer." [72]

### *3.1.2. Risks of AI*

Amodei et al. distinguish **several causes for accidents involving AI systems**:[73] (i) the designer may have wrongly specified the objective function such that maximising it leads to harmful results – even if the system learns. The objective function may ignore crucial aspects of the environment or may fail to reflect the designer's intent; (ii) the designer may fail to carefully evaluate the objective function and the system may produce bad results when extrapolating from limited samples; (iii) the designer may rely on insufficient or poorly curated training data and the system may produce unpredictable bad decisions when given inputs that are different from what was seen during the training phase.

**Bad decisions by AI systems may produce harm in all familiar categories: injuries and property harm, as well as financial harm and other rights, which may be recognised as harm by Member States' tort laws to different degrees.** AI systems or robots acting in the physical space could cause injuries to third parties or harm their property. We could think of autonomous vehicles, planes, or public transport; drones; robots; autonomous traffic management systems; medical devices or precision farming tools. On online platforms, errors in the removal of allegedly infringing online content may cause harm to intellectual property. "Smart" products create the risk of exposure to personal data.[74] AI systems can also cause financial harm, for instance, if algorithms exclude people from insurances. AI systems may also cause harm to other rights, such as the right to privacy or the right not to be discriminated against, which may also lead to financial harm. Overall, harm often depends on the context in which an application is deployed: by whom, for what purpose, and directed at which groups.[75]

## 3.2. Challenges of AI's characteristics for non-contractual liability

For liability, we need to ask not just if AI creates risks, but if AI creates risks that cannot be adequately dealt with under our current liability rules. In other words, do the **unique characteristics of AI** require adapting these liability rules? Given the ambiguity surrounding the concept of AI, we focus on some of the key characteristics of AI identified by the Expert Group Report. The Expert Group Report distinguishes complexity, opacity, openness, autonomy, predictability, data-drivenness, and vulnerability.[76] Given the partial overlap of the risks involved we categories these features under **complexity**, **opacity,** and **autonomy**. To derive policy recommendations for the liability rules that would meaningfully address the risks associated with AI, we assess the relevance of these aspects of AI for the conditions for negligence liability: harm, causation, and fault.

[72] EP Resolution, Annex, Art. 3.
[73] Amodei 2016, p. 1.
[74] BEUC (2020), p. 8.
[75] Response to the Public Consultation of Mozilla, p. 3.
[76] Expert Group Report, p. 7.

### 3.2.1. Complexity

A first characteristic listed in the Expert Group Report and the White Paper is the **complexity** of AI systems. This characteristic is familiar from numerous other products, including IT systems.[77] AI is complex because of the involvement of multiple stakeholders; the interdependence of AI components; and the evolving nature of AI systems.

One aspect of complexity is the **number of stakeholders involved in producing and operating AI systems**. It may be difficult to attribute liability to the manufacturer if a third party was involved, for instance, to supply data. The involvement of multiple stakeholders is neither new nor limited to AI products. Multiple stakeholders are involved in the production of many tangible products, such as cars, which are effectively regulated by the EU's existing liability regime.[78]

Arguably, the involvement of multiple stakeholders may still raise concerns for liability concerning AI systems. Consumers who purchase Internet of Things (IoT) objects, for instance, are confronted with various potential contractual partners, who are in charge of the different services required for the IoT to function properly.[79] These stakeholders may include hardware manufacturers, software designers, equipment and software installers, facility owners, AI owners, and third parties.[80] The various parts of digital goods, such as hardware and digital contents, may be sold separately and produced by multiple parties.[81] Consumers may have difficulty proving why their product does not work, for instance, whether it is due to hardware or digital content. Since consumers carry the burden of proof for the existence of a defect, the costs of determining its cause fall on them.[82]

Secondly, the **components of AI systems are interdependent**: the tangible devices, such as sensors or hardware, interact with the software components and applications, the data itself, the data services, and the connectivity features.[83] As multiple systems become interconnected, the risk of unanticipated or cascading problems grows.[84] Digital goods that depend on data and connectivity may also be more vulnerable to cybersecurity risks.[85] The risk of harm may also increase if there are compatibility problems between components from different manufacturers.[86]

If producers can ensure that components are compatible, the interaction between components may raise problems in litigation settings. When multiple parties are involved, certain actions will be complements and others are substitutes. If an action could be performed by several parties, and something goes wrong, the question may arise who was supposed to do it (see further Section 4.4 below).

Third, **AI systems are said to be complex due to their evolving nature**. If AI applications continue to learn after they are brought into circulation, it is more difficult to assign liability for harm to manufacturers or owners.[87] It is difficult to ascertain if the incorrect output is a "defect" in the meaning of producer liability (see further Section 5.2.2 below).

---

[77] Ebers (2020), p. 44.
[78] Response to the Public Consultation of DigitalEurope, p. 21.
[79] Ebers (2020), p. 44. See also Wendehorst (2016), p. 7.
[80] Benhamou & Ferland (2020), p. 6.
[81] BEUC (2020), p. 5.
[82] Ebers (2020), p. 44.
[83] Benhamou & Ferland (2020), p. 6, referring to European Commission, Liability for emerging digital technologies, Accompanying the document Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, April 25, 2018, SWD/2018/137 final.
[84] Schneier, B. (2018). Click Here to Kill Everyone. *New York Mag.*, <https://nymag.com/intelligencer/2017/01/the-internet-of-things-dangerous-future-bruce-schneier.html>. See also Marcus (2018); Giuffrida (2019), p. 442;
[85] BEUC (2020), p. 6, names the examples of a connected smoke detector failing to detect smoke due to a loss of connectivity, and vulnerability to cyberattacks.
[86] A possible response by manufacturers is to exclude warranties when third-party software is used. However, to encourage competition in the market, it may be beneficial to find a legal solution that encourages manufacturers to allow third-party developers access to their products.
[87] Benhamou & Ferland (2020), p. 6.

### 3.2.2. Opacity

Another challenge for liability lies in the opacity of AI, meaning that AI lacks transparency.[88] This is also referred to as the "**black box-effect" of AI**.[89] Humans, including programmers, have difficulty understanding exactly how input results in the output of AI systems.[90] Injured parties may not be able to identify that they have been harmed or what exactly caused harm.[91] Opaque systems make it more difficult to hold decision-makers accountable or liable for the outcomes of these systems.[92] Opacity may not apply to all forms of AI and may extend to non-AI cases as well.[93]

To alleviate the problems of opaque AI for liability, scholars[94] and policymakers[95] have called for making algorithms explainable, for instance through transparency requirements. Transparent AI systems may create fewer problems for claimants in liability cases and relieve the need for adapting liability rules. However, it is not obvious that transparency would resolve the problems claimants may face in civil litigation. We need to ask what **purpose transparency** is to serve.[96] Two issues may be relevant here: who is meant to benefit from transparency and what information would be needed.

First, we need to ask **who is meant to benefit from transparency**: should this be users, courts, or computer scientists? Users may lack the knowledge to identify a problem of AI products if they arise. However, this does not necessarily pose a problem to liability. For many non-AI products, such as medication, consumers cannot be expected to understand how these work. In civil litigation, expert advice allows courts to make decisions on the technology they do not necessarily understand.

Second, we need to consider **what information about AI systems can realistically be made transparent**. If we understand transparency as tracing back how certain factors affected a specific AI decision, AI transparency could have several dimensions.[97] First, a transparency requirement could focus on the input data. What training data were used for the AI system, and could this data be biased in any way? Second, transparency could aim at allowing observers to verify how input variables affected the output.[98] In the context of a liability case, this would mean asking what factor explains why two similar cases yielded a different decision, for instance in a discrimination case. Third, a transparency requirement could focus on the decisions of AI systems: what was the decision or action not only in a particular instance, but how did the AI system perform, or what did it decide, in a broader set of cases?

In liability contexts, it may be useful for courts to be able to observe how input variables affected the output. However, this may not be feasible for more advanced AI systems. Even if, in some cases,

[88] AI White Paper, p. 17.
[89] AI White Paper, p. 1.
[90] Ebers (2020), p. 48.
[91] Expert Group Report, p. 33: "The more complex emerging digital technologies become, the less those taking advantage of their functions or being exposed to them can comprehend the processes that may have caused harm to themselves or to others. […] It is therefore becoming increasingly difficult for victims to identify such technologies as even a possible source of harm, let alone why they have caused it. Once a victim has successfully claimed damages from a tortfeasor, the tortfeasor may face similar difficulties at the redress level."
[92] Edwards & Veale (2017).
[93] Response to the Public Consultation of DigitalEurope, p. 21: "some of the special issues generally associated with AI, such as the lack of transparency or the unpredictability of concrete individual results, do not apply to all forms of AI but to the more data-driven, probabilistic AI solutions where causality can be more difficult to identify."
[94] See the previous CERRE Report on *Explaining the Black Box*, by De Streel et al (2020); Pasquale (2015); Diakopoulos (2016). However, it has also been shown in a controlled experiment that transparency can increase decision errors by humans who obtain AI-based advice; see Schmidt, Biessmann, and, Teubner (2020); Wachter et al. (2017); Tutt (2017).
[95] European Parliament Committee on Legal Affairs, Civil law rules on robotics (2015/2103 (INL)) 10; Executive Office of the President, Artificial intelligence, automation and the economy (2016); UK House of Lords Artificial Intelligence Committee, AI in the UK: ready, willing and able?; para 105. See also Cath et al. 2018.
[96] Buiten (2019), pp. 15-17; Mittelstadt et al. (2016), p. 6.
[97] Buiten (2019), pp. 14-15. Walt & Vogl (2018) distinguish between the process level, the model level, and the classification level for explainable AI.
[98] Doshi-Velez & Kortz (2017), p. 3; Goodman & Flaxman (2017).

the opacity of AI systems is intentional for reasons of privacy or competitive advantage,[99] often it is inevitable as a result of the complexity of the system. Transparency or explanations are not free:[100] they come at the cost of limiting the complexity and performance of AI.[101] Technological advancement may produce new and better instruments to explain AI decision-making, as the development of "XAI" illustrates.[102] We may also want to encourage this development through the law. To the extent that this is possible, generating explanations for algorithms is likely to be costly, time-consuming, and potentially at a trade-off with the sophistication or accuracy of algorithms.[103] **The implication is that liability rules will have to be suitable to deal with opaque AI systems.**

### *3.2.3. Autonomy*

**AI applications may become increasingly autonomous thanks to the strides made in machine-learning and deep-learning**. This may pose challenges for attributing liability to any party involved under a fault-based liability regime. Combined with opacity, autonomy could make it difficult to trace back specific actions to specific human decisions in their design or their operation.[104]

Applications with a high degree of autonomous decision-making are currently still very rare.[105] Autonomy is not a typical characteristic of applications currently referred to as AI, or of machine-learning systems. The problems related to the autonomy of AI refer to two distinct issues, worthy of more distinction.

**The first aspect of autonomy is the level of control** manufacturers, owners and users have over the actions of AI systems. As machine-learning and deep-learning capabilities advance, AI systems may be technically able to make predictions independently.[106] AI systems may act in ways that humans would not have considered, reducing the control humans have over the outcomes. Scherer explains the example of C-Path, a machine-learning program for detecting cancer, which found indicators for diagnosing breast cancer that contradicted predominant medical thinking.[107] The ability of AI systems to come up with new solutions is among its great benefits. At the same time, it makes it more difficult to attribute a harmful decision of an AI system to the developers. "Unlike traditional engineering and design, the actual functioning of an autonomous artificial agent is not necessarily predictable in the same way as most engineered systems."[108]

**If manufacturers cannot foresee how an AI application will decide or act once placed on the market, it may be difficult to hold them liable.**[109] Product liability holds the manufacturer responsible for the product working as designed, and foreseeing likely problems or harms it may cause.[110] Asaro notes:

> "While there is a degree of unpredictability in the performance of any engineered product, due to failures or unforeseen circumstances of use, there are shared expectations regarding its performance, testing for the limits of that performance and likelihood of failure, and

[99] Kitchin (2017); Mittelstadt et al. (2016).
[100] Doshi-Velez & Kortz (2017), p. 3.
[101] Seseri, R. (2018, June 14th). The Problem with "explainable AI". *TechCrunch*, <https://techcrunch.com/2018/06/14/the-problem-with-explainable-ai/>.
[102] See the previous CERRE Report on *Explaining The Black Box*, by De Streel et al. (2020).
[103] Doshi-Velez & Kortz (2017), p. 3; Reed (2018), p. 6; Datta, Sen & Zick (2016).
[104] EP Resolution, p. 6, number 7.
[105] See also EP Resolution, p. 9, number 24.
[106] Schönberger (2019), p. 193.
[107] Scherer (2016), pp. 363-364.
[108] Asaro (2016), p. 2.
[109] As Sullivan and Schweikart (2019) note in relation to medical devices: "As the AI system becomes more autonomous, fewer parties (ie, clinicians, health care organizations, and AI designers) actually have control over it, and legal standards founded on agency, control, and foreseeability collapse—directly impacting opportunities for recovery of damages based on legal theories of negligence and vicarious liability." Scherer (2016) states: "if the designers of AI cannot foresee how it will act after it is released in the world, how can they be held tortiously liable? And if the legal system absolves designers from liability because AI actions are unforeseeable, then injured patients may be left with fewer opportunities for redress".
[110] Asaro (2016), p. 2.

management of foreseeable risks. In the case of advanced AI, a system that learns from environmental data may act in ways that its designers have no feasible way to foresee"[111]

It may be possible to hold manufacturers liable if the lack of predictability of an AI application is due to insufficient care in developing the application or due to cost-saving measures that lead to insufficient monitoring.[112] The law may need to clarify what that level of care entails for AI products, for instance in safety standards and testing requirements.

**The second aspect of autonomy is the level of automation with which the device operates**. An application with a high level of automation means that little human supervision is required. A system could be autonomous, in the sense that it is difficult to predict its outcomes, but not automate certain decisions, for instance, if it is used to support human decision-making. Automation of decision-making makes it more difficult to ascertain which stakeholder is responsible for the actions of an AI system and what they should (or *could*) have done to intervene.[113]

Often, the degree of automation is considered a gradual scale: as AI gets more advanced, the application will become more autonomous and will operate more safely. However, AI that acts or decides fully autonomously presents different challenges for liability than AI that still requires some level of human supervision. If a task is fully delegated to AI, humans need to be able to rely on it functioning on its own. If humans still need to monitor the system, the human-machine interface needs to work well. Schönberger highlights the example of autonomous driving, which has shown that full autonomy might be safer than requiring human intervention in critical situations.[114] In cases where AI is employed to support human decision-making rather than replacing it, the question arises in which circumstances users are allowed to rely on the AI system and at what point they should override its decisions. **This means that, for safety purposes, automation is not necessarily a continuum: fully autonomous AI may be safer than human action, but an intermediate solution where AI complements human decision-making is only safer than human action if humans supervise it appropriately.**

For instance, the Tesla driver who died in a 2018 crash was playing a video game on his smartphone at the time of the crash, where he should have been monitoring the car.[115] Drivers may rely too much on vehicle automation and fail to concentrate on driving sufficiently.[116]

The tragic crashes of Boeing 737 Max airplanes in 2018 and 2019 are another example of the risk involved in the interaction between humans and technology. The Boeing 737 Max relied on an automated software tool (the Manoeuvring Characteristics Augmentation System, (MCAS)" that was meant to work discreetly in the background. The system was therefore not mentioned in the pilot manual.[117] Investigations identified the MCAS software as the proximate cause of the accidents but illustrated that failing pilot training and regulatory oversight also played roles. [118] The example shows that technology aimed at reducing a primary risk can create or exacerbate a distinctive type of secondary risk, arising from the interaction between the product and the user's experience with it.[119]

[111] Asaro (2016), p. 2 and Asaro (2008).
[112] Asaro (2016), p. 2 notes that "Unpredictability by itself is not an insurmountable problem for liability, insofar as the agents who introduce that unpredictability could be themselves held liable, or the risks from unpredictability could be managed."
[113] According to Chinen (2016): "The more autonomy machines achieve, the more tenuous becomes the strategy of attributing and distributing legal responsibility for their behavior to human beings."
[114] Schönberger (2019), p. 194, referring to Davies, A. (1 January 2017). The Very Human Problem Blocking the Path to Self-driving Cars. *Wired*, <https://www.wired.com/2017/01/human-problem-blocking-path-self-driving-.cars/>.
[115] Rushe, D. (2020, 26th February). Tesla driver who died in 'autopilot' crash was playing on phone, inquiry finds. The Guardian.
[116] As Galasso & Luo (2018a), p. 5, note: "The National Transportation Safety Board determined that the 2016 fatal Tesla crash was partly due to the driver's inattention and over-reliance on vehicle automation despite manufacturer safety warnings."
[117] Palmer (2019), p. 2; Nicas J., Kitroeff N., Gelles D., Glanz J. (1 June 2019). Boeing built deadly assumptions into 737 Max, blind to a late design change. *The NY Times* <https://www.nytimes.com/2019/06/01/business/boeing-737-max-crash.html>.
[118] Palmer (2019), p. 2.
[119] Wendel (2019), 431-432. See also Perrow (2011).

Where AI supports or replaces human decision-making, humans will adjust to this and rely on AI. If the interaction between AI and human supervision may be "non-obvious and difficult to predict",[120] the question for liability is what humans may expect an AI system to deliver. It needs to be clear to what extent a human is at fault for failing to control the device if AI is not fully automated. The liability question may be easier when there is no human-machine interaction, and the AI system functions fully autonomously.

## 3.3. Case studies

Depending on the context, AI may have different implications for the risks involved, as well as if any gaps in our existing liability rules exist. We consider three examples: 1) transportation: Autonomous Vehicles; 2) Healthcare: clinical decision support software; 3) Consumer products: robot vacuum cleaners.

### *3.3.1. Transportation: Autonomous Vehicles (AV)*

The liability for autonomous vehicles has been discussed extensively in the literature over the past years.[121] It is also the field where the EU has the most developed and recent safety rules. This legal framework is composed of three main rules.[122]

The **new General Vehicles Safety Regulation (GVSR)[123] which contains specific definitions and provisions within the field of Autonomous Vehicle (AV)**. This Regulation is the first EU legal instrument defining what are 'automated vehicles'[124] and 'fully automated vehicles'.[125] It contains a set of specific systems that will become mandatory for automated vehicles and fully automated vehicles such as systems that must be able to replace the driver and carry out his tasks or that provide real-time information to the vehicle about its environment (except for the driver availability monitoring systems which do not apply to fully automated vehicles).[126] The regulation also empowers the Commission to adopt delegated acts to specify the technical requirements of these systems.

In addition to the specific provisions related to AV, the GVSR deals with four issues that have potential implications for AV. First, it defines new advanced safety systems such as intelligent speed assistance, advanced driver distraction warning, advanced emergency braking system, and an emergency lane-keeping system.[127] Second, it imposes on manufacturers the obligations to ensure that all vehicles, systems, technical units, and components comply with technical regulatory requirements concerning, *inter alia*, protection against unauthorised use and cyberattacks and remote access to in-vehicle data or software modification that endanger vehicle passengers and other road users.[128] Third, it also requires event data recorder, intelligent speed assistance, and advanced driver distraction warning for all motor vehicles; braking and lane-keeping systems for cars and light commercial vehicles as well as special systems to detect and avoid vulnerable road

[120] Galasso & Luo (2018a), p. 5.
[121] For a literature review see Alawadhi et al. (2020).
[122] For a comprehensive analysis of the EU regulatory framework applicable to Autonomous Vehicles, see EPRS (2021), Cost of Non-Europe report on artificial intelligence in road transport, Annex, Chapter 4.
[123] Regulation 2019/2144 of the European Parliament and of the Council of 27 November 2019 on type-approval requirements for motor vehicles and their trailers, and systems, components and separate technical units intended for such vehicles, as regards their general safety and the protection of vehicle occupants and vulnerable road users, OJ L 325 of 16.12.2019. This Regulation shall apply from 18 July 2022 and replace Regulations 78/2009, 79/2009 and 661/2009. Motor vehicles designed and constructed for transportation of passengers are vehicles of Classes M1, M2 and M3. Motor vehicles designed and constructed for transportation of goods are vehicles of classes N1, N2 and N3. Classes O1, O2 and O3 relate to trailers for motor vehicles.
[124] Article 3(21) GVSR: "a motor vehicle designed and constructed to move autonomously for certain periods of time without continuous driver supervision but in respect of which driver intervention is still expected or required".
[125] Article 3(22) GVSR: "a motor vehicle that has been designed and constructed to move autonomously without any driver supervision".
[126] Article 11 GVSR.
[127] Article 3 GVSR.
[128] Article 4(5) GVSR.

users for buses and trucks;[129] and provides high-level technical requirements for those safety systems, including concerning the processing of personal data.[130]

The **new Approval and Market Surveillance of Vehicles Regulation (AMSVR)**[131] **lays down an administrative type-approval procedure for manufacturers willing to market a vehicle, system, component or separate technical unit in the entire EU territory**. The manufacturer has to demonstrate that each candidate vehicle type, system, component or separate technical unit comply with technical regulatory requirements contained in Annex II AMSVR.[132] EU type-approval certificates are issued by national approval authorities and allow a manufacturer to market vehicles EU-wide without any additional requirements. EU type-approvals are issued after verification of compliance with the relevant requirements. Compliance checks are carried out by technical services designated by approval authorities. During the certification process, manufacturers must establish an information folder and can be required to grant access to any software or algorithm but also, if needed, to provide information or documentation necessary to understand this software or algorithms.[133] Thus, approval authorities and technical services can request information that is necessary to understand software and algorithms underlying the functioning of AV.

National authorities can also refuse to issue EU type-approval certificates for vehicles or components that present high safety risks despite compliance with the relevant requirements.[134] It can happen, for instance, when specific technical requirements do not (yet) exist for components necessary for AV. Moreover, to allow innovation while ensuring safety, the regulation includes a procedure for manufacturers to obtain, under specific cumulative conditions, a type-approval if they use new technologies or new concepts that prevent from complying with the relevant requirements.[135] These type-approvals can only be issued if the manufacturer (i) justifies why new technologies or concepts prevent compliance with the relevant requirements; (ii) ensures a level of safety equivalent to that provided by the relevant requirements, and (iii) provides test results to ensure a similar safety level. After the adoption of implementing acts, the European Commission will decide whether or not to grant an exemption. In the meantime, national authorities can grant provisional exemptions limited to their territories.

**Besides, the AMSVR also contains several provisions that apply during the use of products.** First, each Member State must designate authority for market surveillance to carry checks verifying the compliance of vehicles, systems, components and separate technical units with the requirements of the AMSVR. National authorities can request any information, including access to software and algorithms. National authorities have the power to investigate the compliance of AI-based AV products with the safety requirements of EU law. The AMSVR also empowers the European Commission to carry out checks of compliance with the regulation of the EU market approvals granted to vehicles, systems, components, and separate technical units.

When national authorities grant market approval for any vehicles, systems, components and separate technical units, they must carry out checks to verify that manufacturers produce products that comply with their initial authorisations. These checks are based on products obtained from the

[129] Articles 6 and 7 GVSR.
[130] Article 6 GVSR. Additionally, Recital 10 of the Regulation specifies that advanced emergency braking systems, intelligent speed assistance, emergency lane-keeping systems, driver drowsiness and attention warning, advanced driver distraction warning and reversing detection systems should function without using any biometric information of drivers and passengers.
[131] Regulation 2018/858 of the European Parliament and of the Council of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, OJ L 151 of 14.06.2018. This Regulation applies since 1 September 2020 and repeals Directive 2007/46 of the European Parliament and of the Council of 5 September 2007 establishing a framework for the approval of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles (Framework Directive).
[132] They refer to many United Nations technical regulations on standardisation of car components (e.g. directional equipment, lamps, heating systems).
[133] Article 25(4) AMSVR.
[134] Article 26(5) AMSVR.
[135] Article 39 AMSVR. In 2019 Commission issued a set of guidelines relating to the decision of granting type approval under this procedure. See European Commission (2019), Guidelines on the exemption procedure for the EU approval of automated vehicles.

manufacturers' facilities, and the authority can request access to software, algorithms and any information necessary to understand their functioning. The authorities responsible for types approval must monitor compliance of products with this market approval.

When, based on its checks or notification from type-approval authorities, a market surveillance authority discovers that a vehicle, systems, components and separate technical units present high risks or do not comply with the AMSVR, it must assess the item in question. If the manufacturer fails to remedy the non-compliance or if the risk requires swift measures, national authorities can withdraw or recall the product.

**The Motor vehicles Insurance Directive (MID)** [136] **requires that all vehicles registered in the EU hold mandatory third-party liability insurance to cover civil liability in respect of the use of vehicles**. The MID also ensures that third-party insurance covers physical damages (including to passengers of the car) and damages to property. However, it does not harmonise liability regimes across Member States. The MID establishes mandatory minimum amounts for physical damages (i.e. €1m per victim or €5m per claim) and damages to property (i.e. €1m per claim).[137] The MID establishes a mechanism to simplify and accelerate the settlement of claims and compensation for victims of vehicle accidents.

### *3.3.2. Healthcare: Clinical decision support software*

Healthcare is a field that stands to be revolutionised by AI technologies.[138] AI can improve preventive healthcare by identifying risk factors,[139] reduce health costs by optimising processes[140] and helping to recommend medication and improve life expectancy by supporting the diagnosis of diseases[141] and facilitating complex surgeries.

We focus here on the use of AI applications to support clinical decisions. These applications do not replace the decision of healthcare professionals (HCP) but can help improve their decisions. The use of such applications raises questions about how liability should be assigned between the manufacturer of the AI application, the healthcare professional (HCP), and the hospital.

In some contexts, the AI application may make a recommendation that would harm the patient. For example, a patient may suffer harm due to inappropriate drug recommendations from an AI tool being adopted by the HCP.[142] This raises the question to what extent the HCP may rely on the decision of the AI application. Should the HCP be held liable for following an incorrect recommendation of the AI application? More complex is the case where, without the use of the AI tool, there would have been no attempt to treat the patient. In this case, the wrong drug would not have been given by not using the AI tool. What should be the consequences for liability, and does it matter how serious the consequences would have been from not treating at all?[143] In such cases, the use of AI tools may raise questions for **the causal link** between the HCP's conduct and the

[136] Directive 2009/103 of the European Parliament and of the Council of 16 September 2009 relating to insurance against civil liability in respect of the use of motor vehicles, and the enforcement of the obligation to insure against such liability, OJ L 263 of 07.10.2009.
[137] Note that a Proposal for a Directive amending MID increase these mandatory minimum amounts for physical damages and damages to properties covered by third party civil liability insurance. See Proposal for a Directive of the European Parliament and of the Council amending Directive 2009/103/EC of the European Parliament and the Council of 16 September 2009 relating to insurance against civil liability in respect of the use of motor vehicles, and the enforcement of the obligation to ensure against such liability, COM(2018) 336 of 24.05.2018.
[138] Institute for Public Policy Research (IPPR), Better health and care for all: A 10-point plan for the 2020s The Lord Darzi Review of Health and Care, final report, 15 June 2018 (The Lord Darzi Review, 2018), <https://www.ippr.org/research/publications/better-health-and-care-for-all>.
[139] NHS, GP at Hand, <https://www.gpathand.nhs.uk/>.
[140] As an example, consider Corti, Artificial intelligence that saves lives (<http://www.corti.ai/>); or Peters, A. (11 January 2018). Having a Heart Attack? This AI Helps Emergency Dispatchers Find Out. *Fast Company*, <https://www.fastcompany.com/40515740/having-a-heart-attack-this-ai-helps-emergency-dispatchers-find-out>.
[141] Gulshan et al (2016), Moorfields Eye Hospital NHS Foundation Trust, Moorfields announces research partnership, 3 July 2016, <https://www.moorfields.nhs.uk/news/moorfields-announces-research-partnership>.
[142] Smith & Fotheringham (2020).
[143] Smith & Fotheringham (2020).

resulting harm in the context of liability. If an HCP is liable for adopting AI recommendations if they turn out to have been harmful, this may discourage HCPs from using the AI tool in the first place. This is problematic if the AI tool reduces the overall amount of error.

In the context of diagnosis, the AI application may produce false positives and false negatives (e.g. an AI tool may fail to spot cancer cells, or it may mistakenly identify cells as cancerous when they are not). It appears that human HCPs rarely identify false positives, but they do regularly miss diagnoses. AI tools have the opposite problem: they rarely miss diagnoses, but they eagerly spot anomalous groups of cells that are healthy.[144] The implication is that AI tools are a very useful complement to the skills of HCPs: algorithmic pre-screening can save time and increase the accuracy of diagnosis significantly.[145] A human expert then needs to check the results of the AI tool to eliminate false positives. This, however, means that the increased accuracy comes at a cost: human experts will also spend time double-checking numerous healthy cases. If they do not, this could open them up to liability, for they *could* have been aware of the diagnoses because of the AI tool. To avoid liability, HCPs may need to motivate any deviations from the software's suggestion. This may increase the workload for HCPs if AI tools have a high rate of false positives.[146] As a result, HCPs may be better off not using the AI tool in terms of their liability. In the context of healthcare, and possibly also in other sectors, **it needs to be clarified how liability rules apply for failing to use a proven AI tool. Asymmetry should be avoided: it should not be costlier for HCPs to rely on proven AI tools than to not use them.**

Regulatory safety standards should help ensure that AI systems only enter the market when they improve outcomes. Still, even very sophisticated systems are bound to be imperfect, raising questions of who carries what liability. It could be argued that when the error rate of an AI system is high, some of the liability should shift to manufacturers. However, not all the liability can be shifted to manufacturers and the HCP as the final decision maker should also carry responsibility.

Patients attempting to collect damages from the HCP or the hospital may have difficulty proving that the HCP was at **fault** when AI was involved in the decision.[147] It may also be unfair and inefficient to allocate full responsibility on HCPs if an AI system makes a harmful recommendation or decision. This would disconnect accountability from the locus of control.[148] To establish if an HCP was at fault when an AI system was involved, one needs to ask if it was reasonable to rely on the system in the given situation. If an HCP relied on a certified broadly used AI tool, and the error was not obvious, an HCP may not be at fault for relying on it.[149] Fault in negligence could be determined in conjunction with safety standards and best practices in the medical community: could the HCP reasonably rely on the AI result? Did follow this result blindly or did she check it?

In practice, the inherent opacity of AI systems may make it difficult for injured patients to offer proof of fault and causation.[150] As a result, patients may not be able to successfully claim damages from the HCP, the hospital, or the producer. As was discussed above, providing transparency into the underlying algorithms is unlikely to mitigate these problems completely. **This suggests that reversing the burden of proof for fault and/or causation in the context of healthcare AI devices would be prudent.[151]**

[144] Fry (2018), p. 89 and the references therein.
[145] Fry (2018), p. 90 and the references therein.
[146] Anderson & Torreggiani (2018).
[147] Physicians and hospitals owe a duty of care to their patients, see Smith & Fotheringham (2020).). Anderson et al. note that "Medical negligence is the failure of a medical practitioner to provide proper care and attention that another similarly qualified practitioner would do in a similar circumstance."
[148] Smith and Fotheringham (2020).
[149] See also Schönberger (2019), p. 197.
[150] Schönberger (2019), 197.
[151] The possible scope of such rules is discussed further in Section V below.

### 3.3.3. Consumer products: robot vacuum cleaners

Robots used by consumers in and around their homes are becoming more commonplace. While accidents appear to be rare, they can and do happen. In the case of robot vacuums, for instance, in 2019, a U.S. house burnt down after a robot vacuum clung to a floor heater and caught fire,[152] and another robot vacuum sucked up the tail of its owner's dog.[153] Traditional vacuum cleaners cause accidents and injuries as well,[154] as do many other household appliances. Some robotic household appliances, such as robot lawnmowers, are considerably safer than their traditional counterparts.[155]

Aside from accidents, smart home devices including robot vacuum cleaners may cause data privacy and security harm. Robot vacuums not only collect data about private spaces as they clean, creating a map of the home, they also communicate gathered information into the cloud.[156] This raises privacy issues, for instance, if the data is shared with other companies for marketing purposes (in violation of the General Data Protection Regulation).[157] It may also present security issues: Ullrich et al. identify various possible security breaches with the robot vacuum they tested, including starting and pausing the robot, extracting a map of the victim's apartment, read arbitrary sensor data, leak customer public IP addresses and gain access to the local network of a customer.

For liability purposes, the relevant question is if robot vacuums present new and different risks and if liability rules can adequately address these risks. In case of accidents leading to injuries and property harm, the problem may be a product defect or misuse of the product.

Being subjected to product liability, manufacturers have an interest – and are legally required – to provide precise warnings and safety instructions to consumers, highlighting the limits of an autonomous product. The Roomba manual, for instance, includes in its safety instructions, that "[s]mall children and pets should be supervised while Roomba is cleaning".[158] If the owner of a vacuum cleaner such as the Roomba were to employ the robot in the presence of small children without supervising them, we would argue that the owner acted negligently. In the case of the dog tail, the owner may have ignored the safety instructions for using the robot vacuum. The instructions of a manufacturer on supervising an AI device can act as guidance for the duty of care of the operator.[159] These safety instructions should however be considered in the context of how the product is marketed: if a vacuum cleaner is marketed as a robot vacuum, which frees up consumers' time, consumers should reasonably be able to expect it to function without supervision.[160]

If accidents occur and users did not ignore any of the safety instructions, two issues for liability arise. The first issue is proving the damage. Different from mechanical vacuum cleaners, in the case

[152] Wang, J. (4 December 2019). Robot vacuum causes house fire. *KOB4*, <https://www.kob.com/albuquerque-news/robot-vacuum-causes-house-fire/5570680/>.
[153] Barnes, J. (25 November 2019). Ballwin police rescue dog sucked up by robot vacuum. *KSDK*, <https://www.ksdk.com/article/news/weird/dog-sucked-up-robot-vacuum/63-a8ea9f8a-98d0-4f44-af40-b9a13244d2bc>.
[154] Macgregor (2002) reports on incidences of injuries to young children sustained by contact with a domestic vacuum cleaner.
[155] For instance, traditional lawnmowers cause many injuries. Robotic lawnmowers can significantly reduce these injuries, among other things because their blades automatically stop running if something or someone approaches it. See e.g. Wang, J. (4 December 2019). Robot vacuum causes house fire. *KOB4*, <https://www.kob.com/albuquerque-news/robot-vacuum-causes-house-fire/5570680/>.
[155] Moore, K. (19 May 2016). How Robotic Mowers will save the day in regard to lawn mower accidents. *NKY Tribune*, <https://www.nkytribune.com/2016/05/keven-moore-how-robotic-mowers-will-save-the-day-in-regard-to-lawn-mower-accidents/>. At the same time, robot lawnmowers have been reported to kill animals, see e.g. Parker, S. (26 September 2018). Robot Lawnmowers Are Killing Hedgehogs. *Wired*, < https://www.wired.com/story/robot-lawnmowers-are-killing-hedgehogs/>.
[156] Ullrich et al. (2019).
[157] On privacy issues of vacuum robots see e.g. Astor, M. (25 July 2017). Your Roomba May Be Mapping Your Home, Collecting Data That Could Be Shared. *The NY Times*, <https://www.nytimes.com/2017/07/25/technology/roomba-irobot-data-privacy.html>.
[158] See <https://homesupport.irobot.com/euf/assets/images/faqs/roomba/500/manual/en-US.pdf>.
[159] Marchant & Lindor (2012) argue something similar with respect to partially autonomous vehicles: if the user ignores the manual's warnings about limiting the vehicle's use in certain weather or the driver fails to operate autonomous mode appropriately, he may be found negligent.
[160] As an example, the company iRobot advertises with the ability of their robot vacuums to do the vacuum for their customers. The term "autonomous", however, is not used in relation to the robot vacuums in the advertisements or on iRobot's website.

of robot vacuums, there may not be an obvious malfunctioning indicating that the underlying algorithm was defective.[161] The injured party may have difficulty proving that there was something wrong with the device, what was wrong with it (e.g. the original programming, the subsequent training, or an external factor in the environment), and that it caused the damage.[162]

A second and related issue is that it may be difficult to determine what constitutes a defective robot vacuum. In the house fire case, one could argue that the robot vacuum was defective, as it failed to avoid the floor heater. One could also argue that employing the robot vacuum while the floor heater was active is constituted improper use of the product. If one concludes that the owner was negligent, we need to decide on the boundaries of improper use of the robot vacuum: would it also be negligent to use the robot vacuum around a radiator fixed to the wall? Put differently, the question arises of what level of safety can be required from the robot vacuum. If the vacuum cleaner is marketed as a robot that frees up its owner's time, in what circumstances do we allow it to fail at this promise, and what failure rate do we accept? Liability rules need to specify the scope of producer liability through the concept of defect on the one end, and the scope of fault-based liability of the operator by specifying their duties on the other.[163]

Combining the issues of proving harm and defining what constitutes a defect, injured parties may have difficulty obtaining compensation for their harm. Attributing damage that may be caused by the robot's dynamics to either the producer or the operator may be difficult. **Overall, the main issues for liability of consumer robots such as robot vacuums appear to be proving a defect and, defining a defect, and defining the duties (or fault) on the part of the operator**.

## 3.4. Implications: Gaps in existing liability rules

We identify three possible gaps in the existing liability regime:

- ***Establishing fault***

Much of the existing scholarship examines fully autonomous systems. However, many AI systems are and will continue to be partly autonomous. Complex, semi-autonomous AI systems present a difficult problem for fault-based liability. These systems have the potential to behave in unpredictable ways.[164] This raises the question of **how people who employ AI systems can be said to be at fault when they could not have reasonably anticipated the actions of this system**.[165] Yoshiwaka notes:

> "Given that even the most careful AI programmers are unable to predict or completely prevent highly sophisticated AI injuries without removing AI's autonomy altogether, tort law will not find any person or product at fault and will consequently allocate injury costs to victims."[166]

As a result of the unpredictability of AI systems and the lack of control on the side of users (and even developers, see below), complex automated systems pose unique problems to fault-based regimes. Fault on the side of users would need to be established in terms of a failure to maintain the automated system or to oversee its functioning.[167] Starting from the premise that AI systems are primarily tools, fault-based liability can continue to hold their users to a duty of reasonable care while using it.[168] However, it is not clear whether the decision of a user to put an automated system

[161] Borghetti (2019), p. 67.
[162] See also Steege (2021).
[163] See also Lohmann (2017).
[164] Surden & Williams (2016).
[165] Smart et al. (2017).
[166] Yoshikawa (2019), p. 1165, referring to Karnow (2016), p. 52. See also Selbst (2020), p. 1331 f.
[167] See also Cofone (2018), referring to negligent supervision of an AI system as a possibility, analogous to supervising a child.
[168] See for a US perspective Selbst (2020), p. 1320.

into operation could be considered negligent if the system causes harm, at least not in all Member States.

Arguably, fault-based liability runs into problems particularly for decision-assistance AI, which is designed to interfere with human decision-making. As Selbst notes, "it replaces or augments human decision processes with inscrutable, unintuitive, statistically derived, and often secret code". If these systems are to improve upon human decision-making, and we often lack understanding of *how* it does so, can humans be considered negligent for relying on the AI system, when this leads to harm?

- ***Proving causality***

The **complexity and unpredictability of AI systems may also make it difficult for victims to prove causality**. Developers do not "control" automated systems quite the same way that, for instance, car manufacturers "control" how airbags deploy.[169] This may raise questions of causality – and, in turn, about the division of responsibility between manufacturers and users (i.e. is the harm the result of a product defect or improper use, see below). As Smart et al note:

> "Whereas there often is a relatively traceable and predictable line between design and harm for many potentially harmful non-automated products like band saws (protective covers) and electrical kitchen appliances (short cords), existing software packages for object recognition and control systems are not as well understood and have far fewer default safety mechanisms built-in. […] at least partly attributed to the general unpredictability of the system across broad contexts."[170]

- ***Dividing responsibility***

For advanced AI products, the division of responsibility between manufacturers and operators, and among various manufacturing parties, may not be clear. Automated systems will likely shift responsibility towards manufacturers.[171] The question arises **what the limit of producer liability is for AI systems with a high level of autonomy** – for instance, if any harmful action constitutes a defect, or if we accept that well-functioning AI systems may nevertheless cause harm from time to time. It is not necessarily clear what liability should continue to fall on owners and users.

Moreover, problems may arise when **dividing responsibility among manufacturers and other stakeholders** involved in the functionality of the product, such as data providers (see Section 4 below).

[169] Smart et al (2017), pp. 12-13.
[170] Smart et al (2017), p. 13.
[171] Selbst (2020), p. 1322.

# 04

# EFFICIENT LIABILITY RULES

# 4. Efficient liability rules

Having identified the risks that AI introduces that may have a bearing on liability, we now turn to the question of how to design liability rules to address these problems. This section considers how liability for AI should be designed from a welfare perspective.

## 4.1. Coase theorem and the necessity of liability rules

**The economic purpose of tort liability is to induce injurers to internalise the costs of harm that can occur from their activities, by adjusting their incentives to take precautions to prevent this harm.**[172] When harm occurs outside private agreements, the private costs of the activity to the injured are lower than the social costs of the activity. High transaction costs prevent potential injurers from concluding agreements with all potential victims about harm. Beyond this total welfare goal, liability also allocates economic rents; here distributive goals come into play.

It is useful to recall that in a frictionless world liability rules do not affect the volume of care because all relevant parties have an incentive to reach an agreement that implements the optimal level of care; this is simply an application of the **Coase theorem**. This also applies to third parties, as these parties can also engage with the firm that may harm them and thus generate a negative external effect. Liability assigns property rights and, therefore, affects the surplus of the parties – it can be seen as an instrument of distributional justice.

Consider a self-driving car that hit a bystander. If the bystander were aware of such risk and negotiations between the bystander and the car producer (assuming that the car producer fully controls the risk itself) could engage in efficient ex ante negotiations making sure that the firm applies the welfare-maximising level of care balancing incremental avoidance cost and incremental benefit from harm reduction. Absent liability the bystander would need to compensate the firm for its harm-reducing efforts. By contrast, under strict liability, the bystander would essentially obtain the property right not to suffer any harm and the liability rule kicks in whenever this right is violated fully compensating the bystander when an accident occurs (understanding that financial compensation for physical harm may never be seen as a full compensation). Thus, even in a hypothetical, frictionless world liability rules play a role.

The celebrated Coase theorem however is of little practical relevance when third parties suffer from malfunctioning products or services. There are often many potential victims, and, besides, these potential victims have less information than the producer. Thus, absent liability rules, it seems likely that socially insufficient care is provided by the firm. In the context of AI, risks are certainly diffuse and, for many potential victims, opaque. Thus, absent liability rules (or other interventions) a firm is unlikely to have the incentive to engage in the optimal level of care.

However, it would be wrong to claim that a firm would not have any incentive to engage in any care. This applies in particular to an established firm if the harm it inflicts on third parties is partly internalised by contracting parties. In the example of the self-driving car, buyers or users of such a car may not be indifferent about how likely it is that bystanders are hit. A car with the reputation of e.g. running over dogs may then be avoided by some of these buyers or users. Thus, even absent from liability rules, the firm would have some incentive to reduce the risk to bystanders. This also applies to the internalisation of external effects by other stakeholders, e.g. employees or investors who do not want to be associated with a firm imposing undue risk on third parties. However, this hope for internalizing external effects by stakeholders of the firm has its limits as stakeholders may not care or lack the relevant information. The history of big tobacco tells us that society is not well served when it just relies on the internalization of external effects by stakeholders. (**Product) liability rules and regulations are therefore necessary elements to protect society from socially excessive harm.**

[172] See e.g. Cooter & Ulen (2012), p. 189.

## 4.2. Liability and optimal level of care

**Imposing tort liability on those engaged in activities that may cause harm operates as a mechanism for internalising harmful externalities.** One objective of tort law is to create incentives for potential wrongdoers to invest in safety at an efficient level by making them pay damages.[173] **Tort law should induce potential wrongdoers to take an efficient level of care.** Reaching the efficient level requires compensation at the margin based on expected societal harm at the margin. Transaction costs and their implications for the optimal level of damages should be taken into account. For example, if harmed parties seek compensation only with some probability (e.g. because of lack of awareness or because of the high opportunity cost of going to court), then the efficient compensation to those victims seeking compensation would need to be a multiple of the amount corresponding to the individual harm suffered. For example, if any harmed party suffers the same harm, which we set to 100, and only 60 percent of all those harmed seek compensation, then a harmed party seeking compensation would need to be compensated with 100 / 0.6, which is equal to approximately 167 to implement the efficient level of precaution. Taking precaution generally involves the loss of money, time, or convenience.[174] Therefore, **zero risks are typically not the socially optimal level of risk** since a reduction of risk typically comes at an increasing marginal cost.[175] Depending on the particular application, precaution may take different forms: additional testing of AI-based solutions, possibly by outside experts operating as certifiers, a commitment to human supervision, and a careful design of the HMI interface to reduce human decision-making errors.

**Ethical concerns have been raised against such a cost-benefit approach. Based on ethical concerns it is conceivable to prohibit or at least limit the use of AI for certain types of activities**.[176] However, regarding liability rules, the presumption is that such a prohibition was not taken and that, therefore, society is willing to accept that sometimes harm occurs. Then, society has to quantify the harm e.g. to life and health, and assign monetary values to harm.

Ethical questions may still arise. One is what level of harm from AI is society willing to accept? Following a cost-benefit approach, any autonomous car (marginally) safer than a human driver should be employed. The opportunity costs of not employing AI would be higher than the costs of employing the system.[177] However, to ease trust and acceptance of new technologies, society may want to impose higher standards on AI than on humans.

A related question is **what harms from experimentation or employing imperfect AI today is society willing to accept, to potentially reduce harm considerably in the future**? The most named example in this context is autonomous cars: autonomous driving may prevent countless traffic victims in the future. However, it may also cause fatal accidents, especially in the early stage. The question is how to compare these casualties to those caused by human drivers? This includes the question of how should future costs and benefits be weighed against current costs and benefits? This issue is not specific to AI but is of particular relevance since current experiences feed into future performance.

**One response to ethical reservations is to apply a cost-benefit analysis after specifying some constraints that have to be met**. Another response is to require AI to meet a higher

[173] Cooter & Ulen (2012), p. 190.
[174] Belfield et al. (2020).
[175] Also, additional measures may be less effective. Assuming that precautions reduce the likelihood of an accident or the amount of harm, but at a decreasing rate of success, the optimal expenditure on precautions is finite. The efficient level of precaution prevails when the additional cost of a precaution measure equals the resulting reduction in expected costs of harm ("marginal costs equal marginal benefits").
[176] In some areas, society may be unwilling to replace human decisions with algorithms. See Ebers (2020), p. 50.
[177] In some cases, non-AI approaches could work just as well or better. Some have pointed out that "the goal should not be on employing technology for its own sake.", but that the "focus should instead be on solving a problem and assessing if and how AI could contribute to finding a solution." See Draft Guidelines for Public Procurement of AI published by the World Economic Forum and the UK's Office for AI.

standard than human decision-making and to require that when moving from human decision)making into AI-delegated decision-making, the net benefit or the benefit to a particular group of economic agents (e.g., consumers of a product or bystanders) must increase by a certain amount. However, in the case of rare events that are "unknown unknowns" and, thus, no probability can be assigned to them, it is difficult to put such considerations into practice.

Society may also want to limit the costs of "experimentation" with regulation. For example, AI may be used first in a semi-autonomous way possibly requiring human supervision. This may then provide a controlled learning environment for AI systems. However, it is not clear to what extent such an incremental approach is feasible and, even when it is, whether it is preferable. As discussed above, human supervision creates additional challenges for the HMI interface, as humans must be in a position to interpret and possibly correct AI-based plans of action. In the case of semi-autonomous vehicles, this is at least feasible: humans would be allowed to overrule an AI-based plan of action when parking a car or would be allowed to stop delegating the driving on the highway to an AI-based system and take over control themselves.

Of course, the order of control may also go the other way around, i.e. AI-based systems may overrule human decisions. This issue also appears in the context of semi-autonomous cars and the best solution is likely to be implemented on a case-by-case. For example, AI-based systems may overrule a driver's decision when braking in an emergency. **In general, it is an important decision whether the authority is delegated to a human or an AI-based system[178] and the delegation decision may be reflected by which party (the human or the party responsible for the AI-based system) is ultimately liable.**

## 4.3. Comparing fault-based and strict liability regimes

**Information costs, the role of the injured party, and the value of the (risky) activity are considerations relevant in choosing between a fault-based or a strict liability regime** for a particular activity (or any other liability regime such as strict liability with a rebuttable presumption of harm).

### *4.3.1. Information costs and incentives of the victims*

Under a fault-based regime, the owner of e.g. a drone is held liable if the owner failed to take the safety precautions demanded by the standard of care. The owner is induced to take efficient precautions if lawmakers and courts determine the duty of care correctly. If the standard is set too high or too low, the owner of the drone will be induced to take a suboptimal level of precautions. In the case of AI, a fault-based regime is potentially suboptimal if courts cannot accurately assign liability, because legal conditions for liability, such as fault and causation, are difficult to prove for AI applications. The efficient level of precautions may not be easy to determine for AI on a general level. They may depend on the technical possibilities to control the actions of AI when designing it. There may be a trade-off here between the safety of AI and its sophistication. That is, more sophisticated AI may offer more benefits to users but may also become increasingly complex or unpredictable, and thereby riskier. If owners and users cannot control an AI system, fault-based liability does not serve its goal of steering them towards more cautious behaviour. In other contexts, this has been a reason to introduce a type of risk-based or strict liability.

**The advantage of strict liability is that the legislator or the court does not need to have information on the optimal level of precaution.** A strict liability rule induces the owner of the drone to take optimal precautions because it shifts all the costs of an accident on her. Theoretically (under perfect compensation), a strict liability rule internalises the costs of harm by requiring the injurer to pay for the social costs of his/her activity, regardless of the level of care taken.

[178] Athey, Bryan & Gans (2020).

However, because the injurer bears all the costs, a strict liability **fails to set incentives for victims to take the appropriate care** in situations where they, too, can affect the likelihood of an accident. In the economics literature, this is coined a double moral hazard problem.

### *4.3.2. Level of activity and innovation*

Abstracting from the victims' incentives, shifting the full costs on injurers also means that a **strict liability rule not only induces the optimal level of care but the optimal activity level as well**. If an activity is inherently risky, even despite efficient precautions, we may want to refrain injurers from engaging in this activity altogether (or, at least, to reduce the level of this activity).[179]

**A fault-based regime does not achieve this**, since an injured can avoid paying for the costs of her activity by taking the required level of care. This explains why most jurisdictions impose strict liability for driving a car, for instance. In certain contexts, AI applications could also cause serious harm, even if proper precautions are taken, e.g. because the AI cannot be trained on sufficiently rich data.

However, the flipside of this is that if an activity is beneficial to society, the potential wrongdoer may become too careful. Strict liability may reduce their activity below the efficient level because negative externalities (i.e. harm) are internalised while positive externalities (i.e. external benefits to society) may not all flow back to them. AI applications produce clear benefits to third parties: cars with autonomous features may be safer, AI diagnostic tools may be superior to humans in detecting diseases, and algorithms produce all types of digital services that consumers enjoy. **While employing AI reduces harm as compared to the alternative, there are opportunity costs of not employing AI.**[180]

Moreover, investments in AI applications and their employment may contribute to innovations in AI in other fields as well. A concern for any liability rule and, in particular, strict liability is that start-ups deploying AI may not be able to bear the associated risk and thus go bankrupt, which would shift at least part of the liability to other parties or the injured party if not fully compensated. Furthermore, foreseeing these problems entrepreneurs may not put their efforts into such a start-up in the first place or may not receive funding. Mandatory insurance could, at least partly, address this issue. However, this would come at the cost of negatively affecting the injurer's incentives to efficiently reduce harm and thus prove to be rather costly.

In this context, it should be acknowledged that **liability does not necessarily chill innovation: it may also encourage firms to develop risk-mitigating technologies and improve the design of their products to reduce the likelihood of harm, and in turn, increasing user trust and take-up**.[181] Absent liability, there are often insufficient incentives to do so, and potential users may correctly anticipate such a problem and delay adoption. In other words, liability can be a catalyst of innovation.

[179] Wagner (2019a), p. 30 notes with respect to liability for AI: "shielding businesses from liability for the harm that they cause, for instance, with a view to fostering innovation, also seems problematic. This is not to say that innovation is unimportant or that incentives to innovate should not be generated. It is doubtful, however, whether the liability system is the preferred tool to create such incentives. To shield certain parties from responsibility for the harm that they actually caused amounts to a subsidization of dangerous activities, leading to an oversupply of such activities."
[180] See Belfield et al. (2020).
[181] Galasso & Luo (2018b).

### 4.3.3. Types of risks

To address AI liability, it is useful to elaborate on several economic environments in which third parties experience damage and the effort decision by a party affects the likelihood that damage occurs or the severity of the damage. An important distinction is **whether risks are idiosyncratic across people constituting the third party or highly correlated**.

**In the case of idiosyncratic risk**, from an individual perspective, the damage remains a random event but for the liable party, the outcome is rather predictable. For instance, a firm may invest in reducing the fraction of products that posed a risk to third parties. While an accident is then highly unpredictable for an individual, a firm faces an average number of accidents, which can be predicted rather well. Fault-based liability may be based on calculating an optimal number of accidents (based on a cost-benefit analysis) and the firm may have to contribute to a pool if the number exceeds of observed accidents is above the optimal number, with a payment that is increasing in the number of accidents.

Of course, if the fault is directly observable with little cost for the legal system, damage payments due to fault-based liability can be assessed in individual cases; then the third party receives the full damage in case fault is established. Strict liability would award damages in all cases independent of the level of care. While third parties do better under strict liability, the firm may not have to bear the burden of higher damage payments fully on its own. In particular, when a firm sells a product, it may optimally pass at least part of the increased expected cost per unit under strict relative to fault-based liability through to its customers.

**In many instances, including many cases involving AI risk, individual risks of third parties are highly correlated and a failure is a rare event.** For example, think of insufficient protection of personal data that are hacked despite an AI system that is supposed to detect such threats.

In theory, the application of fault-based liability would work as follows: There is an optimal level of protection implying certain damage in case of failure (damage quantified in Euro, say $X$, and a probability $p$ this damage happens). Keeping the size of the damage constant, we can focus on the failure probability. The optimal failure probability $p^*$ would be compared to the observed failure $p^0$ and damages are awarded such that, from an ex ante point of view, the firm has to pay $(p^0 - p^*)X$. Hence, whenever a failure is observed, it is inferred that the firm's fault increased the risk and, therefore, has to pay $\frac{p^0-p^*}{p^0}X$. For example, the investigation concluded that the optimal failure probability is 1% and the actual failure probability was 5%, then due to fault, with total harm of € 1 million to third parties, damages of € 800k should be awarded as damages. Thus, the idea of fault can in theory be applied to such probabilistic events.

We note that in terms of incentives (taking the presence in the market as given), both liability regimes perform equally well in theory. From an ex ante perspective when deciding about the level of care, a firm minimises the sum of expected damages and avoidance costs, which, under strict liability, is $p^0X + C(1-p^0)$ with respect to $p^0$. This implies that the firm chooses the risk such that $X = C'(1-p^0)$. Under fault-based liability, a firm minimises $(p^0 - p^*)X + C(1-p^0)$ with respect to $p^0$. This implies that the firm chooses the risk such that again $X = C'(1-p^0)$.

The above argument shows that fault-based and strict liability leads to the same (efficient) level of care. The application of fault-based liability leads to practical problems since this requires the court to be able to calculate optimal and actual risk. Strict liability does not suffer from this practical problem, as, in our example, in case of failure, simply € 1 million are awarded. **Therefore, if the individual risk is highly correlated and a failure is a rare event – we may want to call such an environment a high-risk environment – practical considerations make strict liability the preferred option. A downside of strict liability is that the expected payment for an innovator is higher than under fault-based liability.**

**This may lead to socially insufficient innovation if the innovator does not internalise all the social benefits from innovation and therefore refrains from entering the market or scaling up activity.**[182]

## 4.4. Care by multiple parties

**When multiple parties affect the risk of harm, we need to ask who should be targeted by the liability rule. From a welfare perspective, this should be the least cost avoider, i.e., the party which can minimise harm at the lowest cost**. To the extent that some harm-reducing activities are complementary, this may imply that multiple parties should be targeted.

In many AI-based solutions, there are several parties involved in providing a product or service (e.g., a self-driving car, as illustrated above). While damage (e.g., a person hit by a self-driving car) may be easy to show in a court, the question arises which of the involved parties should provide damages and how much the total should be.

To take a look at how different liability rules work out, it is necessary to specify how failures can occur with several parties. We distinguish between two polar environments. In the first environment considered below, care is cumulative; that is the provision of care by one party is a perfect substitute for care provided by another party. In the second environment, care by all parties is essential; that is, the provision of care by one party is a perfect complement to care provided by another party.

Strict liability says that the total damage has to be compensated. As is often the case, it is often unclear which of the parties is to blame. For simplicity, suppose that two parties symmetrically contribute to the risk.

### *4.4.1. Substitute care*

First, consider the substitute case. If at least one of the two parties engages in an effort the risk is assumed to be $p^*$, while if none of the two exerts effort the risk is assumed to be $p^0$. We assume that the socially efficient decision is that one of the two parties exerts effort. If in case of an accident when it cannot be verified which party did not exert effort, the total harm is $X$, and one simple rule would be to equally allocate damages to the parties. With strict liability, each party then would have to pay $X/2$. Such a rule cannot achieve an efficient effort provision.[183] It would be most efficient if the least-cost provider exerts the effort.

If this party can be identified at the outset, one may assign liability to this party. However, this may be difficult to do. Alternatively, the law could specify that a certain type of party will be held liable no matter whether its effort cost is lower than that of other parties. If this party bargains efficiently with the other party, both may agree to shift liability to the least-cost provider. This would guarantee an efficient level of effort at the lowest cost. Similarly, a fault-based liability assigns damages $\frac{p^0-p^*}{p^0}X$ to one party according to a pre-specified rule. As discussed above, in high-risk environments it will be more difficult to implement such a fault-based liability rule.

### *4.4.2. Complement care*

Second, consider the complement case. Here both parties have to exert effort to reduce the risk from $p^0$ to $p^*$. We assume that the socially efficient decision is that both parties 1 and 2 exert effort; i.e., the total cost of effort provision satisfies $C_1(1-p^*)+C_2(1-p^*)<(p^0-p^*)X$. Strict liability that

[182] The chosen liability regime should therefore be seen in the context of public policy towards innovation. The choice of strict instead of negligence-based liability increases the call for public support to innovations to compensate for the higher expected payments to injured parties.

[183] When both parties are equally good at reducing risk, this can be seen as follows. The probability of harm depends on the joint cost the two parties incur, $p(C_1+C_2)$ with $p'<0$ and $p''>0$. The welfare-maximizing solution satisfies $p'(C)X+1=0$. If each party has to pay for half the damage, party 1 minimizes $p(C_1+C_2)X/2+C_1$ with respect to $C_1$ and party 2 minimizes $p(C_1+C_2)X/2+C_2$ with respect to $C_1$. Thus, parties incur costs $C$ with $p'(C)X/2+1=0$. Hence, the overall level of care is less with this solution than in the welfare-maximizing one.

allocates the total harm among the two parties according to some exogenous sharing rule does not necessarily achieve the efficient effort.[184] If the two parties are symmetric, effort provision by the two parties is efficient if $2C(1-p^*) < p^0X$. If each party has to bear half of the damage, a party exerts effort if $C(1-p^*) < (1/2)(p^0-p^*)X$ provided that it expects the other party to exert effort as well. If both parties behave that way, efficient effort is provided.

However, if a party is sceptical about the other party's effort provision if will not exert effort since this does not reduce the probability of an accident. Thus, there may be a **coordination failure**. Coordination failures can be avoided if parties can provide proof of effort that is verifiable in court and if a party that does not provide proof will be held fully liable.

If the effort is not binary (yes/no), but its level can be adjusted, both parties will exert a socially inefficient level of effort. Simply assigning the total damage to the two parties cannot lead to an efficient level of care. The overall payment must be larger than the harm that is inflicted (above the efficient level). The incremental expected payment from not exerting effort must be equal to $(p^0-p^*)X$ for each party; from a legal perspective, this means that **there may need to be punitive damages to implement the socially efficient level of care.**

**The feature of effort being complements may be identified as a particular high-risk environment because the effort of all parties is needed to keep risk at bay**.[185] For example, self-driving cars require reliable sensors and properly functioning AI-based software. If only one of the two has a problem, this is sufficient to significantly increase the probability of harm.

What about assigning strict liability to one pre-specified party? Can this also lead to an efficient level of care? The problem is that the party that is subject to liability may contract with the other party. However, moving part of the liability risk to this other party, it creates a free-riding problem for itself as it is only subject to part of the liability risk. Thus, **assigning liability to one party and efficient contracting cannot resolve the under-provision problem as long as parties only have to cover the harm that has been incurred**. A similar issue arises for fault-based liability rules that only account for the incremental harm beyond the efficient level. It is thus important to acknowledge that in the presence of complementarities in which individual effort cannot be proved in court, merely compensating damages will lead to an inefficient level of care. This holds even under strict liability.

If each of the parties providing care as perfect complements is fully liable for the damage, efficient care will be provided. However, the harmed party will then receive double damages. In the spirit of fault-based liability, **by assigning damages to each party based on the incremental harm above the efficient level, under some conditions, the total payment can then be kept below the money equivalent of the total damage, and still, the incentives for effort provision are efficient**.[186]

To use a numerical example, suppose a lack of care by either one of the two parties implies that an accident occurs with probability p0 = 5%, while with efficient care by both parties this probability is reduced to $p^*$ = 3%. Expected incremental harm from a lack of care is 2% time's damage $X$. When

[184] When both parties are equally efficient in reducing harm given that the other party has contributed more and damage is shared equally between the two parties, no party has an incentive to invest more in reducing the probability of harm than the other party. The problem for party 1 becomes to minimize $p(\min\{C_1, C_2\})\,X/2 + C_1$. For $C_1 \leq C_2$, this gives $p'(C_1)X/2 + 1 = 0$. Thus, the largest effort in harm reduction that can be supported by the behaviour of rational parties satisfies $C_1 = C_2 = C/2$ with $p'(C/2)X/2 + 1 = 0$. By contrast, the welfare-maximizing solution satisfies $p'(C/2)X + 1 = 0$. Hence, under this liability rule both parties spend too little on harm reduction from a welfare point of view.

[185] We acknowledge that complementarity is not specific to AI, see Kremer (1993). At the root of the Boeing 737 Max crashes lies a malfunctioning sensor and its interaction with a software. More precisely, "erroneous AOA sensor reading triggered the plane's automated Maneuvering Characteristics Augmentation System (MCAS) anti-stall software" (Zhang, B. (4 April 2019). Boeing and Ethiopian investigators confirm a faulty sensor was triggered on the 737 Max shortly before it crashed. *Business Insider*,<https://www.businessinsider.com/boeing-ethiopian-investigators-confirm-bad-sensor-triggered-faulty-software-before-crash-2019-4?r=US&IR=T.>).

[186] See Cooter & Porat (2007). From an economics perspective, this is a simple application of the strategic issue in the provision of Cournot complements.

$X$ is 1 million Euros, it is €20k. In case of an accident, each party would be required to pay $((p^0 - p^*)/p^0)X$ = €400k. Thus, the total payment would be €800k, which is less than the total damage of 1 million Euros. As discussed above, the difficulty in applying this idea in practice is the lack of information by the court about $p^0$ and $p^*$.

## 4.5. Liability, regulation, and barriers to entry

Due to liability rules, the party who caused the harm has to make payments to the harmed party. This affects the incentives of the firm and thus the amount of care. However, if accidents are rare and the firm has only a short-term perspective (e.g., because of a high probability to leave the market or financial constraints) it may not fully internalise the expected damages it has to pay.[187] Also, not all third parties may file for liability. The legislator may therefore devise **ex ante regulation** to at least partially deal with the risk to third parties from defective products; this also applies to AI. For example, one may think of a certification procedure for certain types of AI that are applied in sensitive areas (e.g., health).

The question then is whether such certification exempts the firm from liability claims. An important argument against such an exemption is that the firm is typically better informed than the certification agency. Thus, an exemption increases the incentives of the firm to conceal problematic information (this issue arises in several other environments, e.g., for clinical tests of pharmaceuticals). In an AI context, such asymmetric information is likely to be present as well; therefore, **certification requirements that may be introduced for some applications are not a substitute but rather a complement to liability rules**. However, if certification is effective this will lead to fewer damage claims. In this sense, a strict certification regime leads to a less frequent application of liability rules.

An important question is what is the effect of intervention by the legislator (e.g., by specifying liability rules and mandatory certification rules) on market entry? Such interventions often contribute to regulatory barriers to entry. However, they may also increase the trust of other parties and thereby lead to a demand expansion for the affected products or services. This is an important question in the context of AI-based applications since this appears to become an economically important market with many follow-on effects in other industries.

Certification costs are often fixed costs or at least decreasing per unit as the number of units increases that a firm sells. By contrast, the expected damages to be paid under the product liability scale with the volume of activity.[188] This suggests that **certification requirements are more likely to lead to barriers to entry than liability rules**. In both cases, the legislator can create funds to cover damages that are to be paid by small entrant firms or to subsidise the certification process, the former is likely to have undesirable incentive effects. Therefore, a priori there seems to be little reason to devise discriminatory liability rules that apply to large firms only.

However, regulation may well target large firms. This applies when individual risks are highly correlated and societal harm increases more than proportionally in the number of harmed people. Therefore, harm to society can be particularly severe on large social networks and specific regulations may apply to large networks only. By contrast, liability rules directly account for the severity of harm; in our formulation above this would be captured by $X$ increasing with the size of the firm at a growing rate.

[187] In particular, if harm is highly correlated and perceived to be a low-probability event, it may simply exit the market in case a third party is damaged.

[188] We acknowledge that, in both instances, there may be scale economies, as e.g. a larger volume of data can better train an AI-based application and thereby reducing risk. This helps with certification and reducing liability claims.

## 4.6. Implications for liability in the context of AI

The following can be concluded for efficient liability rules for AI-based applications:

- **The efficient level of care**: Tort liability should induce producers and users to take an efficient level of care in designing, testing, and employing AI-based solutions.
- **Endogeneity**: By shifting costs of harm, the rules on liability may influence the design choices of producers in delegating decisions to AI-based systems or humans.
- **Information costs**: The advantage of strict liability, as compared to a fault-based regime, is that legislators and courts do not need to have information on the optimal level of precaution in designing and testing AI-based solutions.
- **Activity level:** By shifting the full costs of harm on injurers, a strict liability rule induces injurers to reduce their level of activity in cases where AI applications are inherently risky, even if proper precautions are taken. A drawback of strict liability is that it may reduce the beneficial use of AI applications below the efficient level, for instance, if their superior performance reduces harm to society as compared to not employing AI.
- **Idiosyncratic risk**: If an individual risk is highly correlated and a failure is a rare event – a high-risk environment – practical considerations make strict liability the preferred option. A drawback of strict liability is that it may lead to socially insufficient innovation if the innovator does not internalise all the social benefits from innovation.
- **Care by multiple parties**: For many AI-based solutions, several parties are involved in providing the product or service. If care by each party is essential to avoid a failure (complementary efforts), and courts cannot verify the source of the failure, even strict liability leads to a socially inefficient level of care when no punitive damages are allowed.
- **Ex ante regulation**: an effective certification regime leads to a less frequent application of liability rules. However, due to the fixed costs on firms, certification requirements are more likely to lead to entry barriers than liability rules.

# 05

# POLICY RECOMMENDATIONS

# 5. Policy recommendations

In respect of liability for AI, the legislator needs to decide on three main issues. First, the **liability rule** for AI needs to be decided. Based on the risks associated with AI discussed in Section 3, and the incentive effects of liability rules analysed in Section 4, this section addresses the three questions for the liability rule laid out at the outset: (i) how responsibility should be divided over actors involved: the recommendations consider the liability of producers on the one hand, and owners or users on the other hand; (ii) what standard of care should apply: the recommendations consider the scope of the strict liability regime currently in place for producers, as well as the possibility of introducing strict liability for owners or users of AI; (iii) what injured parties need to prove: as a possible alternative or complement to the liability standard, the report considers introducing a presumption of harm in the context of producer liability and the liability of owners and users.

Second, the **scope of the liability regime** needs to be decided. The options are introducing a separate regime for all AI, introducing a separate regime for high-risk AI applications or continuing working with current sector-specific rules. Third, the **level of EU harmonisation** needs to be decided. Member States could be allowed to continue applying their national liability regimes; the EU could set a minimum standard or the EU could aim to harmonise liability rules for AI.

In answering each of these questions, the **broader regulatory framework for AI should be acknowledged**. Guiding ethical principles and the regulatory framework of AI will help reduce the risks of harm, by promoting the use of good training data and rigorous testing.[189] They will also improve the effectiveness of the liability regime, in particular transparency and explainability, data, and record-keeping.

In our recommendations, we focus primarily on the efficiency of the rules. We recognise that other goals, such as fairness and ethical considerations, play a role as well. The recommendations are structured as follows. After laying out the broad principles that should guide the liability regime, we first identify to what extent the gaps in our current liability rules can be addressed by updating product liability rules. Second, we consider the need for new liability rules on others, such as "operators", owners or users, and what standard of liability would be appropriate. Next, we consider the possible scope of such a regime. Finally, we discuss the appropriate level of EU harmonisation.

## 5.1. Principles on which the efficient liability regime should be based

Considering the risks associated with AI discussed in Section 3 and based on the efficiency analysis laid out in Section 4, the liability regime for AI should be based on the following main principles:

1. Specify clear liability rules that provide incentives to **all stakeholders** - in particular producers, operators and users - to take an **efficient level of precaution** which, in turn, could facilitate the design of products that minimises risk, as well as the social acceptance and utilisation of new technologies;

2. Place **liability on the least cost avoider**, i.e. the party that can reduce harm at the lowest cost (we have seen that complementarities between parts of AI systems may complicate this); it is also important to place the liability on the party who benefit the most from the use of the new technologies;

3. Be **based on risks of harm**, which may differ depending on the application and the context in which it is used;

[189] AI White Paper, p.18. See also e.g. Galasso & Luo (2018a), p. 6.

4. Ensure an efficient **disclosure of information**, particularly where the asymmetry of information exists between stakeholders;

5. Ensure effective **protection of users** and encourage **innovation** and deployment of AI systems;

6. Balance **proactive** policymaking, anticipating technological changes, with **reactive policymaking**, adapting the rules only after having gained some experience from deploying the technologies;

7. Be **principles-based and flexible**, while allowing for sufficient legal certainty and predictability for all stakeholders;

8. Be **technologically neutral** - the level of protection of users of AI applications should be the same as users of the same application which is not powered by AI;

9. Be **coherent with other EU and national rules**, in particular, the exante rules on safety and surveillance, the national non-contractual and contractual liability rules and rules on insurance;

10. Provide for the **optimal level of harmonisation** at the EU level and respect the principle of subsidiarity.

## 5.2. Liability of producers

### *5.2.1. Rationales for reviewing the Product Liability Directive*

As Section 2.2 discussed, the Product Liability Directive (PLD) attributes strict liability for defective products to producers. A review of this directive is expected at the end of 2021. Given that this review is underway, a likely question to ask is what are the challenges to liability posed by AI that could be resolved by updating the PLD? This question is particularly relevant, given that the scope of revised product liability rules help define how responsibility is divided between manufacturers, owners and users. **The rationale for updating the Product Liability Directive is broader than the concerns identified in relation to AI, but is closely related to technological development**.[190]

AI systems shift the locus of control away from users towards manufacturers.[191] For technical products that do not rely on AI, the manufacturer controls the product's safety features and provides the interfaces between the product and its user, while the user exercises control over the mechanical device when employing it in real-world situations.[192] For AI systems, users will be able to exert much less control. As a result, accidents will become less dependent on the care taken by the individual user. The liability of the user is likely to increasingly recede into the background, meaning that the role of liability of the manufacturer becomes more significant for injured parties to obtain compensation.[193] **In short, where producers are in a better position than consumers to control risk, an incentive-based approach would shift the relative burden of liability towards producers**.[194] This incentivises producers to reduce the AI system's risk through designing and manufacturing the system.

[190] ELI Guiding Principles for Updating the Product Liability Directive for the Digital Age ("**ELI Guiding Principles**"), January 2021. The European Law Institute Guiding Principles for Updating the Product Liability for the Digital Age ("ELI Guiding Principles") also note, "[t]he rapid development of digital technology and the integration of physical goods with the digital sphere" and call for a review of the Product Liability Directive.
[191] Wagner (2019a), p. 37.
[192] Wagner (2019b), p. 602.
[193] Seehafer & Kohler (2020), p. 213, Lutter (2017), p. 281, Wagner (2019b), p. 602.
[194] Galasso & Luo (2018a), p. 5; Lohmann (2016), p. 338.

The Commission has identified three problems with applying the provisions of the PLD in the context of IoT and autonomous connected systems. The first is the complicated product or service value chain, with interdependencies between suppliers, manufacturers and other third parties. The second is the uncertainty in relation to the legal nature of IoT devices, i.e. whether they are products, services, or products that come with the sale of a service. The third is the autonomous nature of these technologies.[195] The Expert Group Report and the White Paper on AI both concluded that some key concepts in the PLD require clarification to be apt to deal with emerging digital technologies.[196] The European Parliament has also called on the Commission "to review the Directive and consider adapting such concepts as 'product' 'damage' and 'defect' as well as adapting the rules governing the burden of proof".[197]

In the following, we consider the possible ways in which central concepts of the PLD could be updated to reflect today's consumer products. At the outset, two important observations should be made. First, the PLD has a horizontal scope and here, we consider only the reasons for reviewing it in light of risks posed by digital products in general and AI systems in particular. Second, the PLD aimed to facilitate damage actions by the victims by creating a strict liability regime while safeguarding the interests of the manufacturers by setting limits in the application of the strict liability. It is key that such original balance within the PLD not be upset by its revision. Third, any extension of the PLD scope will also automatically increase the scope of EU harmonisation of liability rules, which involves a specific trade-off as explained below. In the following, the possibilities to review the notions of "product", "producer", "defect", the burden of proof and the available defences are considered.

### *5.2.2. Product and software*

One of the first and most discussed issues regarding the PLD review is whether software should be covered in the notion of "product". **Currently, tangibility is a central aspect for determining whether the PLD applies**. Hardware components of an AI system would certainly be deemed a product,[198] as would software integrated into tangible products.[199] However, if the software and the hardware originate from different companies, the treatment of software as a product determines if the manufacturer of the software could be held liable next to the hardware manufacturer.[200] For standalone software, the medium becomes decisive: where software is stored on a tangible medium, such as a DVD or a Flash-drive, it qualifies as a product.[201] However, if the software is downloaded, the application of the PLD is unclear.[202] Member States have moreover applied the concept of software from the PLD differently in their national implementations.[203]

**In the age of digitalisation, differentiations between tangible and intangible objects of use may be more difficult to justify**.[204] It is unclear why the mode in which computer programs are stored, copied, and distributed should be relevant for the application of the PLD. Digital content is increasingly replacing the functions that physical objects performed at the time of the entry into

[195] Communication from the Commission of 10 January 2017, Building a European data economy, COM(2017)9.
[196] AI Commission Report; Expert Group report (2019), pp.27-28.
[197] European Parliament Resolution of 12 February 2020 on automated decision-making processes: ensuring consumer protection and free movement of goods and services (2019/2915(RSP)).
[198] Allain (2013); Navas (2020), p. 167 w.r.t. robots.
[199] Navas (2020), p. 167, referring to Fairgrieve et al. (2016), p. 47. Case law and jurisprudence has largely already taken this approach. Given that the Directive covers electricity, it could be argued that a product does need to be tangible, see Ebers (2020), p. 58.
[200] See further Stöber/Pieronczyk/Möller 2020 612
[201] Written Question No 706/88 by Mr Gijs de Vries to the Commission: Product liability for computer programs, Official Journal (OJ) C 114, 8.5.1989, 42.
[202] See further Lutter (2017), 282. Some authors take the position that the PLD already now extends to digital content, e.g. Koch (2019), p. 106, Wagner (2017), pp. 717-8 and Spindler (2011), pp. 41-43.
[203] See e.g. Nemeth & Carvalho (2019), p. 160 on the differences between the German and the Austrian implementation.
[204] See also e.g. Stöber, Pieronczyk & Möller (2020), p. 613. With respect to healthcare, Sullivan and Schweikart (2019) note that: "The legal reasoning of not allowing products liability to extend to software is that software, as opposed to hardware, is "technology that helps healthcare providers make decisions by providing them with information or analysis" and that the final decision of care rests with the health care professional, while "blatant hardware defects" would instead be subject to products liability suit against the manufacturer".

force of the PLD.[205] Software is no longer distributed on tangible storage devices such as hard drives, CDs, or USB sticks. Instead, it is downloaded from a cloud server and no tangible asset is ever exchanged.[206] The main purpose of the PLD was to ensure a fair distribution of the risks associated with industrially manufactured between the injured party and the manufacturer.[207] The risks associated with downloaded software do not appear very different from their traditional counterparts supplied on CDs.[208] Once the software is introduced to a computer, it brings about material and tangible changes.[209] This is obvious where software is integrated into a machine[210] but is also easily imaginable for intangible software: one could think of an insulin therapy app used by a patient making an error,[211] or malware corrupting all of a consumer's files. The risks involved in software, irrespective of its medium, therefore support including software in the notion of products.[212]

Such an approach would raise questions on **how to delineate products from services**. Items that were once consumed as products purchased by the consumer are delivered not only in the cloud but often also as services by a service provider.[213] For instance, where consumers would previously buy a CD, they now have a subscription to Spotify. Digital goods have blurred the distinction between products and services.[214] As cloud-computing abilities improve, more AI systems may be operated on service model as well – not just digital goods, but physical ones as well.[215] As a result, **it may become increasingly difficult to draw a sharp line between products and services for IoT and AI systems**. It has been argued that the distinction between products and services is less justified with respect to many digital goods, given that their risks may well be the same.[216] It has been proposed that in the medium to long term, either a common liability regime will have to be adopted for both, or clear definitional criteria will need to be developed.[217]

### *5.2.3. Producer*

Physical products are often supplied in connection with digital content or a digital service.[218] As Section 4.4 illustrated, for many AI-based solutions, several parties are involved in providing the product or service. If care by each party is essential to avoid a failure (complementary efforts), and courts cannot verify the source of the failure, the level of care will be inefficiently low.[219] While the producer of the end product is finally responsible, the boundaries of responsible parties can become blurred when AI systems process data provided by third parties or when they autonomously collect data from the environment, controlled by user-specific settings.[220] Responsibilities may become blurred particularly if products are unbundled, and original equipment manufacturers lose control over the safety features of the products they put into circulation.[221] For IoT devices as well, vulnerabilities may be based on the lack of hardware protection, software failures or both.

[205] Expert Group Report, p. 43. See also Seehafer & Kohler (2020).
[206] Wagner (2019b), p. 604.
[207] Product Liability Directive, Recital paras. 2 and 7.
[208] Wagner (2019b), p. 604.
[209] Alheit (2001).
[210] Alheit (2001), in footnote 107.
[211] Seehafer & Kohler (2020), p. 214 name the example of an error in an insulin therapy app causing a patient to suffer life-threatening hypo- or hyperglycaemic lapses.
[212] The view that the PLD should apply independent of the mode in which computer programs are stored, copied and distributed is shared by the Expert Group Report, Schmon 2018, 254, Stöber, Pieronczyk & Möller (2020), p. 613, Steege (2021), p. 7, Wagner (2019a), p. 42, Wuyts (2014), p. 6, Weber (2017), p. 210.
[213] Rachum-Twaig (2020), p. 1157 and the text in footnote 93.
[214] BEUC (2020), p. 7; Expert Group Report, p. 28.
[215] Rachum-Twaig (2020), p. 1172 names example of robots.
[216] Benhamou & Ferland (2020), p. 13; EU Report, 43; BEUC (2020), p.13.
[217] Marcus (2018).
[218] ELI Guiding Principles, p. 6.
[219] See also Steege (2021), p. 12.
[220] Seehafer & Kohler (2020), p. 216.
[221] Wagner (2019a), 50-51; Wagner (2019b), p. 607.

This makes attributing liability, once a breach or failure has occurred, a long and complex process, requiring specialists.[222]

It may be **useful to clarify in what way AI developers, algorithm trainers, data collectors, controllers, and processors, and manufacturers of the devices incorporating AI software[223] are pulled into product liability**. The overriding importance of training data for the capabilities and functioning of AI systems supports clarifying the role of data providers in product liability.[224] If new categories of producers are defined, as the EP Resolution proposes with the concept of the "backend operator",[225] this would need to be defined very clearly for producers to know where the boundary is drawn.

### *5.2.4. Defect*

Product liability decisively depends on whether the product is defective. The definition of "defect" is therefore pivotal in determining producer liability for autonomously operating systems.[226] Two aspects of "defect" need to be clarified in the context of AI systems (and for digital products more broadly): the expectations consumers are entitled to have of AI products and the meaning of "defect" in the context of autonomous decision-making.

#### *5.2.4.1. Safety expectations*

**The notion of "defect" relates directly to the safety expectations consumers are entitled to have of the product**. The PLD defines the moment a product was brought into circulation as decisive for producer liability. Producers are not liable under PLD for a defect arising after a product was placed on the market, reflecting that they have no control over the product from that moment onwards. However, if product safety relies on a producer's updates to the software, the lack of control argument no longer applies.[227] The same holds for AI systems that are intended to continue learning once they are placed on the market.[228] **A reform could consider the dynamic nature of software products, IoT devices and AI systems**.[229] One option could be to extend liability to producers that fail to provide updates relevant to the safety of the product.[230] It may be useful to clarify if consumers may expect these updates to be delivered throughout the life-cycle of the product.[231]

Such a product monitoring obligation is alien to the current PLD.[232] To prevent an over-broad and open-ended liability, if such an approach were followed, clear criteria would be required. First, it must be clarified how long such an obligation should reasonably exist.[233] Second, failure by users to install the software update should count against liability.[234] The Commission has already pointed out

---

[222] Weber (2017).
[223] Giuffrida (2019), p. 442.
[224] Etzkorn (2020), p. 364.
[225] EP Resolution, Article 3 under (f).
[226] Von Westphalen (2020), p. 549; Borghetti (2019), pp. 63 ff., Seehafer & Kohler (2020), pp. 23 ff.; Wagner (2017), pp. 724 ff.
[227] Expert Group Report, p. 28; BEUC (2019), (2020).
[228] Commission AI Report, p. 15. See also Seehafer & Kohler (2020), p. 214.
[229] See Expert Group Report, p. 43; Benhamou & Ferland (2020), p. 13; BEUC (2020), noting that "the Digital Content Directive already provides that the trader (i.e. the seller) shall ensure that the consumer is informed of and supplied with updates, including security updates, that are necessary to keep the digital content or digital service in conformity for the period of time (Art.8 2b). Similarly, the Directive on the sales of goods (2019/771) also provides that a seller is liable for digital elements being in conformity with the product including for updates provided for as a long as the consumer may reasonably expect (Art.7.3)."
[230] Nash (2021), p. 7-8 notes: "There is no objective requirement for the producer to develop patches for vulnerabilities, unless this has been agreed in the sales contract. Reference is made to 'security updates' in the Directive, but this term is not defined." The Expert Group Report notes that "the EU has confirmed in Directive (EU) 2019/771 on the sale of goods that a seller is also liable for such digital elements being in conformity with the contract, including for updates provided for as long a period as the consumer may reasonably expect, and Directive (EU 2019/770 establishes a similar regime for digital content and digital services. The proposed features of a producer's strict liability […] follow very much the same logic, though on different grounds."
[231] Schmon (2018), Seehafer & Kohler (2020).
[232] Seehafer & Kohler (2020), p. 217.
[233] Steege (2021), p. 12.
[234] Seehafer & Kohler (2020), p. 217; Steege (2021), p. 12.

that subsequent updates cannot be the sole responsibility of the manufacturer: the user would have the obligation to install safety-relevant updates.[235] For AI systems with learning capabilities, the question is whether the producer should be liable for defects that develop as the system learns. One issue is whether the defect can be traced back to a defect in the algorithm itself, a related question is what monitoring duties the producer has with respect to learning capabilities.[236]

#### *5.2.4.2. Notion of defect*

As AI systems become more autonomous, the question arises whether any instance of harm constitutes a defect, or whether it should be accepted that a well-functioning AI system could still cause harm.[237] If ***some*** failure rate is accepted, the question arises of ***what*** failure rate is acceptable. It needs to be clarified how far the concept of defect extends for deliberate, but undesirable operations of AI systems with self-learning capacities. For sophisticated AI systems, it may not be possible to draw the line between harm resulting from AI's autonomous decisions and harm resulting from a defect.[238] We consider two options: first, extending the concept of "defect" for fully autonomous AI applications to any harm they cause or, second, distinguishing more clearly between different types of defects;

**Extending the concept of the defect to any event of injury** has been justified by the fact that the producer designs the learning process for the AI system, is best-placed to judge whether the product is safe enough to be put on the market, and profits from selling it.[239] Such a rule would moreover encourage producers to inform users about contexts in which the application is unable to work fully autonomously. However, **we identify several drawbacks to this solution**. From a practical perspective, we can expect producers not to market applications as fully autonomous anymore.[240] Producers would likely add extensive product manuals outlining the contexts in which users still have a duty to monitor the system. More generally, the **consequences of the liability rules on the design and marketing of AI systems should be given serious thought**. If products with a higher autonomy level are treated differently under product liability, this will likely affect how products are marketed and/or how they are designed.

Aside from this practical problem, it would be unreasonable to require absolute safety in the context of liability.[241] Certain situations, such as in healthcare, may require demanding absolute safety because of the high stakes involved,[242] which is reflected by regulatory safety standards. Generally, extending strict liability to AI manufacturers so that they are responsible for any AI harm shifts an undue portion of the burden on manufacturers.[243] Such a regime would force AI manufacturers to bear the negative externalities without compensation for the value of the tremendous positive externalities of AI.[244] Moreover, waiting for nearly perfect AI before employing it may be more costly than accepting a failure rate, which should be reflected in the liability rules.[245] At the same time, shifting full responsibility on manufacturers would place a too little burden on the owners and users

[235] AI Commission Report, p. 15. See also Seehafer & Kohler (2020), p. 217.
[236] Seehafer & Kohler (2020), p. 214, referring to Hey (2019).
[237] Borghetti (2019), p. 61 notes that "The mere fact that a product, including an algorithm, caused harm or damage does not make it defective.".
[238] Benhamou & Ferland (2020), p. 7; EPRS 2019, p. 28: "With regard to non-sophisticated robotics and AI, the proof of a regular material defect, such as a machine defect, deficient safety systems or malfunctioning, would not constitute problems. However, with regard to future sophisticated robotics and AI, it will become more burdensome to prove such defect, particularly given the self-learning ability of these products and the asymmetric information between the producers and consumers, due to which it will be difficult to ascertain what exactly caused damage, as well as the capability of autonomous behaviour. Lawful autonomous behaviour of robotics and AI causing damage may be considered not a defect."
[239] See also Cauffman (2018), p. 530.
[240] As Lemley & Casey (2019), p. 1327 note: "calling for the total elimination of the danger is tantamount to calling for a prohibition on a product or service itself".
[241] See e.g. Schrader (2016), p. 243 and Von Westphalen (2020), p. 250.
[242] Lutter (2017), p. 283
[243] Yoshikawa (2019), pp. 1165 and 1171.
[244] See Section 4. See also Yoshikawa (2019).
[245] RAND, Kalra & Groves (2017).

of AI systems, who benefit from employing AI and impose risks on others by doing so (see further below).

A second option would be to clarify the different types of defects and potentially differentiate the applicable liability regime. The PLD uses a single criterion to establish the defectiveness of a product. It does not distinguish between different types of defects, as is the case e.g. in American product liability law.[246] One possibility could be to **limit strict liability to manufacturing defects, while a presumption of fault could be applied for defects in design and instructions to users**.[247] In practice, however, this would likely limit strict liability in the context of AI systems as compared to other products. For AI systems, defects are more likely to originate in their design and instructions than in their manufacturing.[248] Instead, the notion of design defect as a type of product defect could be clarified for the context of AI systems. If we accept that AI systems that are free from software bugs, hardware errors, or failures of engineering precaution will nevertheless harm others,[249] we need to consider what failure rate is acceptable. Rather than focusing on an individual AI system, we need to ask what error rate in a fleet of AI systems that operates by the same algorithm constitutes a defect.[250]

Generally, the safety requirements placed on the manufacturer increase with the risks associated with the product.[251] We may also expect AI systems to be safer than the "dumb" products they are replacing. However, we need to consider **how much safer than human decision-making we require AI systems to be**. It may not be possible or useful to compare the performance of an AI system with how a carefully acting human would have behaved in a specific situation. The first reason is that precisely because we require AI to do better than humans a **comparison with reasonable human decision-making** is pointless.[252] Second, the **point of reference differs**: in the case of a human being, it is the decision to act in an individual case, while for an AI system, it is whether the programming for an entire series of products could and should have been done more careful to prevent the occurrence of the damage.[253] Courts would need to identify shortcomings that could have been avoided by alternative programming.[254] Self-learning AI systems that originally function well and develop a malfunction in practical use could be considered already initially not error-free.[255] A third reason is that the **pool of accidents** that an autonomous system causes may be easily avoidable by humans – one can think of the ability of an autonomous car to recognise a white truck in a bright environment. Despite making errors that humans would not, AI systems may overall still make significantly fewer errors. As a result, it may be misguided to compare the standard for safety to humans.[256]

Overall, **with regard to autonomous AI systems, we need to consider what design flaws for AI are unacceptable and what error rate is unacceptable**. Moreover, the burden of proof (discussed below) and regulatory safety standards may help mitigate the challenges that autonomous AI systems pose for the concept of the defect.

[246] See further Wuyts (2014), p. 10.
[247] Navas (2020), p. 168.
[248] Navas (2020), p. 168; Hubbard (2014), pp. 1821-1823; Ebers (2017), pp. 111-112. Nash (2021), p. 8 describes that for software more broadly, a critical difference with ordinary products is that it is impossible to distinguish between faulty products, part of a batch in which a defect arose in the manufacturing process, and faultless products. Vulnerabilities in software would be inherent in the code that affects all the software, meaning that all defects would be considered a defect in design, as opposed to a defect in a given device.
[249] Lemley & Casey (2019), pp. 1327-8.
[250] Wagner (2019b), p. 606, points out that a drawback of this approach for competition may be that "this method would lead to a finding of all the algorithms in the market as defective except for the safest of them all." See also Steege (2021), p. 10.
[251] See Lutter (2017), p. 283, and on the German situation e.g. Schrader (2016), p. 243 and Von Westphalen (2020), p. 250, and the references therein.
[252] Borghetti (2019), p. 69.
[253] Seehafer & Kohler (2020), p. 214 and footnote 23 therein.
[254] Wagner (2019b).
[255] See (Etzkorn) 2020, p. 362; Zech (2019), p. 196.
[256] Wagner (2019b), p. 605.

Regulatory safety standards can also help reduce negative unintended consequences from autonomous decision-making. **It is important to increase our understanding of how AI systems learn once they are placed on the market, to take the appropriate regulatory steps.** A "defect" becomes more difficult to recognise or even define if AI devices continue to learn on their own once they are on the market. Such devices would be less predictable and harder to control. If AI devices are thoroughly tested after a learning process and "frozen" when placed in the market, harm from unintended actions may be less likely. If regulation precludes AI products from entering the market without "**freezing**" them this could reduce the need for interpreting "defect" more broadly in the product liability rules. However, such an option should always be weighed against the lost benefits of not employing these systems with learning capabilities.[257]

### *5.2.5. Burden of proof*

The PLD requires injured parties to prove that the product was defective and that it caused the injury. This "is not necessarily problematic, depending on the criterion to determine causality and the standard of proof used to determine the defect."[258] Outside the AI context, proving the defect may pose difficulties for the consumer because of "the technical complexity of certain products, the high cost of expert evidence, the parties' unequal access to information (particularly about the production process) and the fact that some products are not retrievable after they have been used".[259] National courts have developed ways to facilitate the burden of proof in such situations, including by disclosure obligations for the producer, or by allocating the costs of experts' opinions.[260]

**For AI products, proving a defect may nevertheless be difficult**,[261] given that the uncertainty about what constitutes a defect of an advanced AI system. For instance, if an AI diagnosis tool delivers a wrong diagnosis, "there is no obvious malfunctioning that could be the basis for a presumption that the algorithm was defective".[262] Depending on the definition of a defect, users may be asked to show that harm was the result of a flaw in the AI device, and not of its autonomous decision-making. Causality is governed by national rules, given that the PLD does not define a causal relationship.[263] Proving causality in the context of AI harm may be difficult, especially if some human supervision was still required. The injured party may have difficulty showing that the AI system, not his negligence, caused the harm.[264] AI developers may also try to argue that it is impossible to anticipate precisely how AI systems will act, meaning that the harm was unforeseeable.[265] While this is unlikely to succeed as a defence, such questions could arise when AI did exactly what it was intended to do (act autonomously) and nevertheless caused harm. The assessment of the causal link will often require expert advice, the cost of which may discourage injured parties from suing.[266]

Reversing the burden of proof has been proposed to facilitate claims for parties injured by highly complex technologies.[267] However, this would significantly alter the current distribution of risks to the detriment of the manufacturers.[268] It would also depart sharply from the current principles of

[257] Kowert (2017) notes: "One of the primary benefits of artificial intelligence is its ability to learn and mold itself with new experiences, resulting in it taking on almost human characteristics. Without allowing it to continue to do so, artificial intelligence is relatively useless."
[258] Cauffman (2018), p. 530.
[259] Wuyts (2014), p. 24.
[260] Wuyts (2014), p. 24 and the references therein.
[261] Barfield (2018).
[262] Borghetti (2019), p. 67.
[263] See on different national approaches e.g. Wuyts (2014), p. 25.
[264] Kowert (2017), P. 184 notes on causation: "When reliance on the Tesla autopilot system resulted in a fatal crash in June of 2016, Tesla quickly tried to shield itself from liability by pointing out that the negligent interactions of the driver were a more immediate cause of the crash, than the actions of the programmer.", The Tesla Team, A Tragic Loss, TESLA BLOG (June 30, 2016)
[265] Kowert (2017), p. 191.
[266] Cauffman (2018), p. 530.
[267] De Meeus (2019), p. 151; Wagner (2017), p. 747; Seehafer & Kohler (2020), p. 216; de Bruin (2016), p. 495, who acknowledges that this could negatively impact innovation.
[268] Koch (2019), p. 110 notes that this would effectively be nothing less than reallocating the overall risk in disguise.

the PLD.[269] Given that AI systems may be equipped with event logging or recording systems, victims may moreover be able to get access to better data about the cause of an accident than they used to.[270] A better alternative to facilitate the burden of proof for the injured party could be the **introduction of a lower standard of proof. This could be accompanied by evidence disclosure duties**,[271] cost-shifting rules for expert advice, or – as was as data protection rules permit – requirements to collect data about the functioning of the system, allowing them to retrace possible causes for an error at a later stage.[272]

#### *5.2.6. Defences*

A review of the PLD should also consider the scope of the defences available to producers, particularly the development-risk defence.[273] AI systems give this defence more relevance: if an AI system with learning capabilities causes harm, the producer may be able to argue that the particular learning and decision-making process was not foreseeable.[274] One could argue that consumers should not bear the risk of gaps in the knowledge about the safety of new technologies.[275] At the same time, the defences are included to maintain incentives to innovate, the requirements for excluding liability are high, and the burden of proof lies with the producer.[276] In light of AI systems, it needs to be assessed what is the 'state of scientific and technical knowledge in relation to machines powered by automated decision making. In a review of the PLD, **clarifying the defence could be justified**.[277]

### 5.3. Liability of operators

Unless we follow the broadest view of "defect" under product liability, AI systems will inevitably cause harm that cannot be recovered from the producer. In such cases, the question arises under what conditions can victims collect damages from the producer or someone else under the general liability rules in Member States? This section maps the possibilities for harmonising liability rules for users or owners of AI. It considers three questions: who could be held liable next to the producer; the standard of care that should apply to these parties; the scope of a possible harmonised regime.

#### *5.3.1. Standard of care for operators*

While AI systems shift the locus of control to producers, producers do not influence the final use of the AI system. It is therefore justified to attribute some liability to the party who owns the AI-powered product (owner) or who uses it (keeper): the "operator".[278]

Following the conclusions from Section 4, there are several **reasons to keep operators of AI systems accountable**. First, much of today's AI technology is not fully autonomous and requires at least some level of human supervision. In these situations, it is important to maintain liability for operators, encouraging them to take precautions in supervising the AI system.[279] Second, even for

[269] See further Seehafer & Kohler 2020, p. 216.
[270] Koch (2019), p. 110, Spindler (2015), p. 772.
[271] Schmon (2018), p. 6.
[272] Von Ungern-Sternberg (2018), p. 9.
[273] Benhamou & Ferland (2020), pp. 13-14. See further ELI Guiding Principles, pp. 10-11.
[274] Seehafer & Kohler (2020), p. 215.
[275] De Meeus (2019), p. 152, referring to Machnikowski (2016), notes that the producer receives the benefits of the distribution of innovative products, while it is the consumer who bears the risk associated with the lack of knowledge available regarding that technology.
[276] Schrader (2016), p. 243 and the references therein.
[277] Schmon (2018). On the defence see further Etzkorn (2020), p. 363.
[278] The Expert Group Report defines as operator as "the person who is in control of the risk connected with the operation of emerging digital technologies and who benefits from their operation". Janal (2020), p. 199 e.f. defines as "the party that is responsible for running the autonomous system", providing "the data necessary to run the system, which oversees possible machine-learning processes and which initiates the necessary update pushes for the software". As Janal acknowledges, often this party will be the producer of the system. It does not need be, however, as is e.g. the case with software for a computer. Under the first definition, the operator is closer to the owner, keeper or user, under the latter it resembles the producer. The EP Resolution captures both definitions. It distinguishes between the 'frontend operator' – the person who controls the risk connected with operating the AI-system and benefits from its operation –and the 'backend operator' – the person who defines the features of the technology and provides data and backend support service. See EP Resolution, art.3 (e) and (f).
[279] See also Galasso & Luo (2018a), p. 6.

highly autonomous AI systems, the operator decides if and how to employ it. Liability provides an incentive for operators to keep an AI device updated and ensure that it is used properly.[280] Third, the operator benefits from employing AI. Holding producers liable for every case of harm, even those they have no control over and are not capturing the benefits from, may harm beneficial innovation.[281]

#### *5.3.1.1. Fault-based liability and control as a criterion*

These reasons support holding **operators liable at least for those aspects of employing AI that they can control**. For autonomous AI systems, the operator of the system does not influence its "behaviour".[282] As a result, "the tortious duties of care or traffic duties are limited to ensuring the safe use of the robot." [283] In most Member States, a standard fault-based liability regime would apply for users of products of any kind, including AI systems.[284] Under standard fault-based liability, the operator is liable where the harmful conduct of the AI system is due to her negligent behaviour within this sphere of control.

The **drawback of a fault-based regime is that courts would need to set the optimal level of care, which may be difficult in the field of AI** where technology and its applications advance continuously.[285] The types of harm themselves may also be unpredictable and new in nature.[286] Overall, harm may not be foreseeable when advanced AI systems are involved.[287] To establish fault, the information provided by the manufacturer could be informative, as is discussed in detail in Section 5.3.3 below. "The autonomous system may only be used by its operator following its intended use, whereby the manufacturer's information on the safe use of the product must be observed. If the operator violates the manufacturer's specifications by misusing or abusing the product, he is liable for any damage caused to third parties."[288] This could also be clarified contractually: Tesla, for instance, requires its buyers to sign a contract that mandates they agree to keep their hands on the wheel at all times, even when the autopilot is engaged.[289]

Nevertheless, especially in the Member States that rely primarily on fault-based liability, such as Germany, it may be difficult to establish that the operator was at fault when an AI system she employs causes harm. From a compensation perspective, following the control criterion may also be considered undesirable, given that it would propagate limiting the liability of operators as compared to liability for ordinary products.

#### *5.3.1.2. Strict liability and the human-AI relation*

Another approach is to attribute operators' extra-contractual responsibility for the behaviour of their AI system depending on the typology of the human-AI relation.[290] Humans will be held responsible for the autonomous actions of AI systems. It would not be the first time legal systems provide for the responsibility and agency of another entity, think of corporations; the difference would mainly be that, other than corporations, AI systems cannot be reduced "to an aggregation of human beings as the only relevant source of their action".[291]

Several suggestions for legal analogies for AI have been made. A first is the parent-child relationship: One could allow operators to evade responsibility only if they can prove it was not possible to prevent

[280] See also Galasso & Luo (2018a).
[281] Benhamou & Ferland (2020).
[282] Wagner (2020), 790, Lemle & Casey (2019), p. 1315
[283] Wagner (2020), p. 790 (translated).
[284] Wagner 2019b, p. 606.
[285] Rachum-Twaig (2020), p. 1144.
[286] Rachum-Twaig (2020), p. 1149 ff.
[287] See also Bartneck et al. (2021), p. 42-43.
[288] Wagner (2020), p. 790 (translated).
[289] Kowert (2017), p. 203.
[290] Pagallo (2012), p. 56.
[291] Pagallo (2012), p. 45.

a machine's action.[292] This would follow the approach of Member States for the liability of parents for their children.

A second option is the principal-agent relationship: In employing AI systems, operators impose risks on others. If we replace the AI system by its human principle for identifying the liable party, the principal is properly incentivised to prevent damages and to invest in achieving an optimal level of activity.[293] Particularly where a corporation operates an AI system, we may think of the corporation as operating the robot on its behalf.[294]

A third possible analogy is the owner-animal relationship: In their erratic and unpredictable behaviour, AI systems resemble animals.[295] The liability standard for animals varies across the Member States, and within some Member States, e.g. depending on the purpose of the animal/AI system.[296] German law differentiates liability according to the degree of danger emanating from the animal, as well as the benefit of the use of the dangerous animal to society.[297]

**The parent-child relationship and the owner-animal relationship as examples of special liability regimes reflect the not fully foreseeable or controllable behaviour of independently acting beings. In this sense, they may serve as a blueprint for the liability of operators for AI systems they employ**. Under this view, "a new generation of robots induce novel types of human responsibility for others' actions" besides liability for the behaviour of children, pets, and employees.[298] At the same time, these examples also illustrate the existence of abnormal danger, even if diligent care is taken. Strict liability is usually justified with the consideration that a particular danger emanates from certain useful and therefore permitted facilities or activities. Those persons who are served by the facility or activity should also be assigned the disadvantages caused.[299] Neither argument applies in the context of AI systems, which often promise a significant increase in safety as compared to their non-AI or human counterparts.[300] For this reason, AI systems may be of important value to society and imposing strict liability on their operators may impose an excessive burden.[301]

**In the case of the principal-agent relationship, the key consideration is that parties can shift risk to another party.** This argument may hold in the context of AI systems. However, a challenge would be that in an employment context, we still require that the principal has control over the agent, which may be difficult to establish for unpredictable behaviour by an AI system.[302]

These examples illustrate that finding an appropriate analogy for AI systems in existing rules on strict liability is not straightforward. At the same time, Member States differ considerably in what contexts they subject people to strict liability. As was illustrated in Section 2.2, France attributes strict liability for any "keeper" of a "thing". However, other Member States only depart from the standard fault principle in isolated cases. This means that an **EU approach introducing strict liability for AI would constitute a sharp departure from the standard liability regime in several Member States**.

[292] Pagallo (2012), p. 56.
[293] Lior (2020), p. 4.
[294] Rachum-Twaig (2010), p. 1151.
[295] Lior (2020), p. 18.
[296] See with respect to Germany e.g. Borges (2018), pp. 981-2.
[297] Borges (2018), p. 981
[298] Pagallo (2012), 45
[299] Pehm (2018), p. 260, referring to Koziol et al. (2014), p. 7 f.
[300] Wagner (2020), p. 791; Rachum-Twaig (2020), p. 1158.
[301] Rachum-Twaig (2020), p. 1145.
[302] Rachum-Twaig (2020), p. 1151.

### 5.3.2. Level of EU harmonisation

There are two main arguments in favour of EU harmonisation of the liability regime for AI-based applications. First, it helps ensure the same level of protection for all users in Europe. Second, under a harmonised regime, operators face the same rule throughout Europe, ensuring a level playing field. Conversely, the following two factors support national diversity. First, the diversity of rules between Member States allows for experimentation and learning, which may be particularly beneficial at the beginning of the deployment of a new category of technologies.[303] Second, choosing not to harmonise the rules also preserves the coherence of the national liability systems. Considering this, **a limited form of harmonisation is justified, in particular where there is a set of EU harmonised safety rules that EU harmonised liability rules may usefully complement. It is questionable whether we need a harmonised regime on the European level for AI in general**.

On the one hand, the existing differences in the liability regimes of Member States reflect different approaches and preferences to attributing costs of accidents. Harmonising rules for specific sets of products and activities at the European level cut through the internal coherence of these systems. On the other hand, it is unlikely for the EU liability framework to provide a truly unified set of rules. The liability rules will still need to be interpreted by national courts, according to various national procedural rules that affect liability. Indeed, experiences with the e-commerce Directive[304] as well as with the antitrust private damages Directive[305] illustrate that European harmonisation in the area of liability can only go so far: claimants still face national civil procedural rules that affect the outcome of the case. Against this background, it needs to be further analysed if the benefits of a horizontal regime for AI are large enough to justify disrupting the national liability regimes.

In the following, several options for harmonisation are discussed: a harmonised regime for AI in general, a harmonised regime for high-risk AI and special liability rules for high-risk AI following existing sector-specific regulation. In each case, first, the possible scope of such a regime is considered, followed by the possibilities for the standard of liability.

### 5.3.3. Baseline standard

#### 5.3.3.1. Scope: identifying AI systems

**Introducing a harmonised operator liability regime for AI systems may lead to delimitation difficulties**. It may be difficult to delimit the scope of application of a general liability regime for autonomous systems or AI systems. It may be impossible to find "a clear-cut and at the same time general criterion for distinguishing between "ordinary" and "autonomous" machines."[306] Questions may also arise as to whether "any product containing artificial intelligence be covered, regardless of whether it was the cause of the damage or not".[307] A general AI liability regime may also not be justifiable, given that a specific liability regime is required "in relation to liability for autonomous systems to the extent that they give rise to the risk of damage caused by unforeseeable behaviour".[308]

#### 5.3.3.2. Liability standard for AI systems

***If*** rules are to be harmonised at the European level, three broad options are possible for the baseline regime applicable to ***all*** AI applications. Option 1 consists of introducing a fault-based regime with

[303] Five years ago, Reed et al. (2016) noted that "it is far too early to devise a liability regime for machine learning generally, because in the current state of development of the technology the law would rapidly fail to accord with technological change."
[304] Directive 2000/31 of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce), OJ [2000] L 178/1.
[305] Directive 2014/104 of the European Parliament and of the Council of 26 November 2014 on certain rules governing actions for damages under national law for infringements of the competition law provisions of the Member States and of the European Union, O.J. [2014] L 349/1.
[306] Borges (2018), p. 981 (translated).
[307] Steege (2021), p. 13.
[308] Borges (2018), pp. 981-982.

a higher duty of care. Option 2 consists of introducing a fault-based regime with a rebuttable presumption of the fault and/or causality link. This is the option suggested by the European Parliament in its Resolution of October 2020.[309] Option 3 consists in introducing strict liability (which is discussed below for high-risk AI applications).

**Under options 1 and 2, operators of AI devices should have to comply with a duty of care in choosing to employ an AI system, maintaining the system, and supervising it**.[310] The duties of the operator necessarily relate to the level of autonomy of the AI device. If the operator may reasonably expect the AI device to act fully autonomously, the operator has no duty to monitor an AI device.[311] **The responsibility of the operator decreases and that of the producer increases, if a product promises to function fully autonomously when used for its intended purpose**. The duty of the operator is then limited to using and maintaining the device properly.

To allow operators to uphold their monitoring and maintenance duties, **producers should be obliged to instruct operators properly on the use of the product**. If producers face a strict liability standard, they should have an interest in providing precise warnings and instructions to buyers to avoid producer liability. If products do not operate autonomously in all circumstances, we can expect producers to issue warnings urging users to monitor the device. Manuals for vacuum robots, for instance, include extensive safety instructions.[312] If the owner of a vacuum robot were to employ the device in the presence of small children without supervising them, the owner acted negligently. More generally, the instructions of a producer on supervising an AI device can act as guidance for the duty of care of the operator. **A liability rule that shifts liability to producers for fully autonomous AI devices would promote information disclosure by producers**.

However, information disclosure can have drawbacks: endless lists of warnings are likely to be ignored by consumers, in the same way that general term and conditions are not read.[313] To ensure that consumers are effectively informed about their devices, setting standards for the information supplied to consumers may still be desirable. A possible solution to this problem would be to **regulate information duties or to make information more accessible by introducing "autonomy labels for AI"**, akin to the European energy labels. The autonomy labels could be aligned with certification processes and other safety regulations and would indicate to consumers what level of supervision is required when using an AI application. Given that the autonomy labels would provide information about the delegation of decisions to the AI system and the humans involved, these labels can inform courts when assigning liability to producers, operators and users. A drawback of this could be that producers may be discouraged from developing AI systems with increased autonomy if this increases their liability, even if this would be a safer alternative to "semi-autonomous" systems that still require human oversight in crucial situations. Nevertheless, autonomy labels could help resolve information problems of courts, by setting clear standards for the division of responsibility for harm involving AI systems.

### *5.3.4. The stricter standard for high-risk AI applications*

#### *5.3.4.1. Scope: identifying high-risk AI applications*

If a stricter standard is to be introduced for certain AI applications, we distinguish two possibilities. **The first option consists of aligning the scope of the stricter standard with existing sector-specific rules while the second option amounts to introducing a horizontal liability framework for newly pre-defined high-risk AI applications.**

[309] See EP Resolution, Art. 8.
[310] Expert Group Report, p. 44. Janal (2020), p. 193 notes that, "the users of an autonomous system may be held liable for the acts of the system if they have breached a duty of care, particularly in operating and supervising the autonomous system."
[311] See also Janal (2020), p. 193.
[312] See https://prod-help-content.care.irobotapi.com/files/s_Series/s9/ownersGuide/ownersGuide_enUS.pdf.
[313] See the previous CERRE Report on *Smarter Consumer Protection Rules for the Digital Society*, by De Streel & Sibony, October 2017.

The first option has the advantage of limiting the risks shifting from the victims to the operators as it will limit the strict liability to the sector already defined in the law. Such an option also ensures coherence with safety rules which are also defined at the sector level. Finally, from a legitimacy point of view, this option ensures that the scope of strict liability rule is defined by the legislator when adopting sector regulation and not by the Courts when interpreting criteria to define high-risk applications. Conversely, the second Option has the advantage of being more flexible and adaptable for technologies that evolve quickly. This second option is also favoured by the European Parliament in its Resolution of October 2020.[314] The Resolution recommends that all high-risk AI-systems be exhaustively listed in an Annex to Regulation it proposes.

As **AI applications differ in both the benefits and the risks they create for society, it is appropriate to differentiate in the regulatory and liability requirements that apply to different AI applications. However, we identify three potential problems with the second option introducing a horizontal liability framework for high-risk AI applications**.

First, **listing "high risk" AI applications may presuppose that AI applications create similar risks regardless of the context in which they are applied**. AI encompasses various technologies, which may be used in a wide range of applications, which in turn could be employed in various contexts. One of us has argued elsewhere that it would be preferable to target regulation on the concrete contexts in which AI is applied, rather than setting standards for AI more broadly.[315]

Second, **existing sector-specific regulation already reflects the need to differentiate regulation according to the context in which technology is applied**. Introducing a horizontal liability regime for high-risk AI would constitute a departure from the existing EU approach to liability. For various reasons, it may be better to streamline liability rules for high-risk AI with existing sector-specific regulation.[316] One reason is that including liability rules in sector-specific regulation would limit the interference of these rules with the coherence of Member States' national liability regimes, as discussed above. It would allow the Member States to maintain their general rules on strict liability, which would be complemented by enhanced obligations in certain sectors. A horizontal liability regime for a limited number of applications would cut through this system.

Third, under the second option, the key question is **how to define the scope of such a regime**. It needs to be sufficiently clear for users (or other types of operators), and courts to understand what applications are covered by such a framework. Indeed if a strict liability framework was to be introduced for high-risk AI systems, cases are likely to revolve around the question of whether a specific device is AI, and whether it is a high risk,[317] as this would determine whether it is covered by the general fault-based or by strict liability. To avoid introducing a new source of legal uncertainty, a European regime would need to clearly define which AI applications or activities are covered by it. A strict liability regime would **need to define not only the technologies that are high-risk but potentially also the applications or contexts that it covers**. The risk of a certain technology may be different depending on the context. Therefore, the risk would have to be determined for a certain device and, in the case of general-purpose devices, for each particular use of that device.

According to the Commission White Paper, high-risk applications should meet the following two cumulative criteria: (i) first, the AI application is employed in a sector where, given the characteristics of the activities typically undertaken, significant risks can be expected to occur such as healthcare, transport, energy or parts of the public sector and (ii) second, the AI application in such sector is used in such a manner that significant risks are likely to arise; the assessment of the level of risk of a given user could be based on the impact on the affected parties such as legal or similarly significant effects for the rights of an individual or a company, risk of injury, death or significant material or immaterial damage; effects that cannot reasonably be avoided by individuals

[314] See EP Resolution, art.4.
[315] Buiten (2019).
[316] Buiten (2019); Reed (2018).
[317] See also Lohsse, Schulze & Staudenmayer (2019), p. 21.

or legal entities. [318] The Commission White Paper also notes that a definition of AI would need to be "sufficiently flexible to accommodate technical progress while being precise enough to provide the necessary legal certainty". The same holds for the definition of "high risk".

The European Parliament defines high risk as 'a significant potential in an autonomously operating AI-system to cause harm or damage to one or more persons in a manner that is random and goes beyond what can reasonably be expected; the significance of the potential depends on the interplay between the severity of possible harm or damage, the degree of autonomy of decision-making, the likelihood that the risk materializes and the manner and the context in which the AI-system is being used.[319] The European Parliament Resolution proposes to list all high-risk AI-systems in the EU legislation that should be reviewed at least every six months. In practice, maintaining such a list may be burdensome. The risk of each AI application may need to be reviewed regularly: after it is placed on the market, an AI system may evolve and new vulnerabilities may arise. Risk assessment would then need to be repeated once a product is already placed on the market.

Fourth, the question is **how many AI applications could ultimately be covered by a uniform horizontal regime**. On the one end, some applications are already covered by specific regulation, such as autonomous vehicles or medical devices. These AI applications can continue to be governed by sector-specific rules. For some AI-driven applications, we may find it useful to extend the existing regime for vehicles or other forms of transportations. One could think of drones. For such devices, compulsory liability insurance schemes may also need to be imposed, similarly as for car owners.[320] On the other end, there is a large group of AI systems that is not high-risk. For this group, strict liability would not be justified if we broadly follow the existing principles for strict liability in the Member States. Accordingly, an EU horizontal strict liability regime would cover the group of AI systems that ***should*** be covered by stricter liability rules but ***would not*** be covered by any (extended) sector-specific policy. If this group is small, introducing a horizontal EU liability regime may not be justified.

Given the need to differentiate for AI-driven applications based on the context in which they are used, the question is how broadly a horizontal regime can ultimately apply. **If most high-risk AI applications will be regulated elsewhere, following the sector-specific rules with a liability regime would be the preferred approach.**

#### *5.3.4.2. Liability standard for high-risk applications*

If a baseline regime applying to high-risk AI applications is to be introduced, three broad options are possible: Option 1 with fault-based regime with a higher duty of care, option 2 with fault-based regime with a rebuttable presumption of the fault and/or causality link and option 3 with strict liability.

**Option 1 would require the legislator to establish a duty of care on operators that are sufficiently clear to be uniformly interpreted throughout the EU, and that allows victims to establish that the operator was at fault**. The difficulty is to establish what duty of care operators owe to others when they employ a (semi-) autonomous AI system. Establishing fault is easily established when operators use an AI system to deliberately cause harm, but it is much more difficult for unintended harm.[321] For decision-assistance technology relying on AI, a different standard may need to be found: the difficulty there is that it is designed to (partly) replace human decision-making. The standard of care would need to clarify in which cases operators can rely on the technology, and when doing so would constitute a fault. Thus such an option does not relieve victims from the problem of proving fault and causality and likely leads to different interpretations from courts throughout the EU.

[318] AI White Paper, p. 17.
[319] EP Resolution, Annex, Art. 3(c).
[320] See e.g. Borges (2019), Navas (2020), p. 166, Levy (2020).
[321] See also Von Ungern-Sternberg (2018), p. 6: Kowert (2017).

**Option 2 has the advantage of accommodating victims with a rebuttable presumption of the fault and/or the causality link**. From the perspective of victims, this may be the preferred solution (even to strict liability, particularly if a presumption of causality would be included). It may help to establish causality where AI systems display autonomy and the operator has little control.[322] The law would still need to specify the duty of care on the operator, to supervise or monitor the AI system.[323] As AI systems reach higher levels of autonomy, such a duty may become more difficult to establish in concrete cases, as discussed above. At some point, AI systems are arguably no longer tools used by humans, but rather machines deployed by humans that act independently of direct human instruction.[324]

**The risk profile of some AI systems may justify attributing strict liability to operators, following option 3**. Following the considerations laid out above, three arguments would support imposing strict liability on operators of a selected group of AI systems, for instance in sector regulation. First, the advantage of strict liability to establishing a "duty to supervise" under fault-based liability is that it ensures compensation for victims also in cases where, even if operators monitor an AI system, they may not be able to prevent harm if an AI system acts in completely unexpected ways.[325] In cases where the risks associated with an AI system are high, the "abnormal danger" argument mentioned above could justify decoupling liability from fault, rather than raising the standard of care or reversing the burden of proof. Second, strict liability can save the high transaction costs that injured parties would need to expend to litigate liability issues involving autonomous systems where the fault is difficult to establish.[326] Third, a strict liability regime may be more predictable. It would likely lead to fewer interpretation variations across national courts in Member States.

[322] See also Lior (2020), p. 14.
[323] For a proposal (in the U.S. context) see Rachum-Twaig (2020), pp. 1168-1170.
[324] Vladeck (2014), p. 121.
[325] Janal (2020), p. 199.
[326] Cf. Vladeck (2014), pp. 146-7 with respect to autonomous vehicles. See also Spindler (2018), p. 50; Lohmann (2017), p. 169.

# REFERENCES

# References


Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press.

Alawadhi, M., Almazrouie, J., Kamil, M., & Khalil, K. A. (2020). Review and analysis of the importance of autonomous vehicles liability: a systematic literature review. *International Journal of System Assurance Engineering and Management*, 11(6), 1227-1249.

Alheit, K. (2001). The applicability of the EU product liability directive to software. *Comparative and International Law Journal of Southern Africa*, 34(2), 188-209.

Allain, J. S. (2013). From Jeopardy! to jaundice: the medical liability implications of Dr. Watson and other artificial intelligence systems. *Louisiana Law Review*, 73(4), 7.

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2), 211-36.

Alter, A. (2017). *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked*. Penguin Books.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*.

Anderson, T., & Torreggiani, W. C. (2018). The impact of the introduction of artificial intelligence to Ireland. *Irish medical journal*, *111*(8), 799.

Asaro, P. M. (2016). The Liability Problem for Autonomous Artificial Agents. *AAAI Spring Symposia*, 190-194.

Asaro, P. M. Peter (2008). From Mechanisms of Adaptation to Intelligence Amplifiers: The Philosophy of W. Ross Ashby. In M. Wheeler, P. Husbands & O. Holland (eds.), *The Mechanical Mind in History*, Cambridge, MA: MIT Press, 149-184.

Athey, S. C., Bryan, K. A., & Gans, J. S. (2020). The Allocation of Decision Authority to Human and Artificial Intelligence. *AEA Papers and Proceedings*, 110, 80-84.

Barfield, W. (2018). Liability for autonomous and artificially intelligent robots. *Paladyn, Journal of Behavioral Robotics*, *9*(1), 193-203.

Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2021). *An introduction to ethics in robotics and AI*. Springer Nature.

Belfield, H., Hernández-Orallo, J., Ó hÉigeartaigh, S., Maas, M. M., Hagerty, A., and Whittlestone, J. (2020). *Consultation on the White Paper on AI: a European approach*. Report by the Centre for the Study of Existential Risk. Available at <https://www.cser.ac.uk/news/response-european-commissions-consultation-ai/>.

Benhamou, Y., & Ferland, J. (2020). Artificial Intelligence & Damages: Assessing Liability And Calculating The Damages. In P. D'Agostino, C. Piovesan & A. Gaon (eds.), *Leading Legal Disruption: Artificial Intelligence and a Toolkit for Lawyers and the Law*. Thomson Reuters Canada.

Bertolini, A. (2020). *Artificial Intelligence and Civil Liability: Legal Affairs*. Study for the JURI Committee by the Directorate-General for Internal Policies, Policy Department for Citizens' Rights and Constitutional Policies.

BEUC (2019). AI RIGHTS FOR CONSUMERS. Available at <www.beuc.eu/publications/beuc-x-2019-063_ai_rights_for_consumers.pdf>.

BEUC (2020). The European Consumer Organisation, PRODUCT LIABILITY 2.0 How to make EU rules fit for consumers in the digital age. Available at <www.beuc.eu/publications/beuc-x-2020-024_product_liability_position_paper.pdf>.

Bibal, A., Lognoul, M., de Streel, A., & Frénay, B. (2020). Legal requirements on explainability in machine learning. *Artificial Intelligence and Law*, 1-21.

Boeglin, J. (2015). The costs of self-driving cars: reconciling freedom and privacy with tort liability in autonomous vehicle regulation. *Yale Journal of Law & Technology*, *17*, 171.

Borges, G. (2018). Rechtliche Rahmenbedingungen für autonome Systeme. Neue Juristische Wochenschrift, 977-982.

Borges, G. (2019). Product Liability 2.0 – Mere Update or New version? In S. Lohsse, R. Schulze, & D. Staudenmayer (eds.), *Liability for Artificial Intelligence and the Internet of Things*, Nomos, 145-164.

Borghetti, J.-S. (2019). How can Artificial Intelligence be Defective? In S. Lohsse, R. Schulze, & D. Staudenmayer (eds.), *Liability for Artificial Intelligence and the Internet of Things*, Nomos, 63-76.

Buiten, M. C. (2019). Towards intelligent regulation of Artificial Intelligence. *European Journal of Risk Regulation*, 10(1), 41-59.

Calo, R. (201). Robotics and the Lessons of the Cyberlaw. *California Law Review*, 103(3), 513-63.

Carpenter, B.M. & Collins, C.G. (2012). The Shirt Off My Back: Using the Relationship Between a Product and a Service to Your Advantage, International Association of Defense Counsel Newsletter. Available at <www.bakerdonelson.com>.

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': the US, EU, and UK approach. *Science and engineering ethics*, *24*(2), 505-528.

Cauffman, C. (2018). Robo-liability: The European Union in search of the best way to deal with liability for damage caused by artificial intelligence. *Maastricht Journal of European and Comparative Law,* 527-532.

Chinen, M. A. (2016). The co-evolution of autonomous machines and legal responsibility. *Virginia Journal of Law & Technology*, 20(2), 338-93.

Chung, J., & Zink A. (2018). Hey Watson—Can I Sue You for Malpractice? Examining the Liability of Artificial Intelligence in Medicine. *Asia Pacific Journal Health Law & Ethics*, 11(2), 51-80.

Cofone, I. N. (2018). Servers and Waiters: What Matters in the Law of AI. *Stanford Technology Law Review*, *21*, 167.

Cooter, R. B., & Porat, A. (2007). Total Liability for Excessive Harm. *Journal of Legal Studies*, 36(1), 63-80.

Cooter, R. B., & Ulen, T. (2012). *Law and Economics* (6th edition). Pearson.

Datta, A., Sen, S., & Zick, Y. (2016). Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems. *2016 IEE Symposium on Security and Privacy*, 598-617.

de Bruin, R. (2016). Autonomous Intelligent Cars on the European Intersection of Liability and Privacy. *European Journal of Risk Regulation*, 7, 485-501.

De Meeus, C. (2019). The Product Liability Directive at the Age of the Digital Industrial Revolution: Fit for Innovation? Journal of European Consumer and Market Law 8(4), 149-154.

De Streel, A. & Sibony, A.L. (2017). Towards Smarter Consumer Protection Rules for the Digital Society. CERRE Policy Report.

De Streel, A., Bibal, A., Frenay, B., & and Lognoul, M. (2020). Explaining The Black Box: When Law Controls AI. CERRE Policy Report.

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56–62.

Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books.

Doshi-Velez, F., Kortz, M. et al. (2017). Accountability of AI under the law: The role of explanation. *arXiv preprint arXiv:1711.01134*.

Ebers, M. (2017). Autonomes Fahren: Produkt- und Produzentenhaftung. In B.H. Oppermann and H. Stender-Vorwachs (eds.), *Autonomes Fahren*, CH Beck, 93-126.

Ebers, M. (2020). Regulating AI and Robotics. In M. Ebers & S. Navas (eds.), *Algorithms and Law* Cambridge University Press, 37-99.

Edwards, L., & Veale, M. (2017). Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke Law & Technology Review*, *16*, 18-84.

EPRS (2021), Cost of Non-Europe report on artificial intelligence in road transport. Available at < https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2021)654212 >.

Etzkorn, P. (2020). Bedeutung der» Entwicklungslücke «bei selbstlernenden Systemen–Rechtliche Fragen zur fortdauernden Softwareentwicklung durch maschinelles Lernen im Praxiseinsatz. *Multimedia und Recht*, 6, 360-365.

European Law Institute, Twigg-Flesner, C. (2021). Guiding Principles for Updating the Product Liability Directive for the Digital Age, *ELI Innovation Paper Series*. Available at <https://europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ELI_Guiding_Principles_for_Updating_the_PLD_for_the_Digital_Age.pdf>.

Fairgrieve, D. (2019). Product Liability in the United Kingdom. *Journal of European Consumer and Market Law*, 8(4), 170-72.

Fairgrieve, D., Howells, G., Møgelvang-Hansen, P., Straetmans, G., Verhoeven, D., Machnikowski, P., A. Janssen & Schulze, R. (2016). Product liability directive. In P. Machnikowski (ed.), *European Product Liability: An analysis of the state of the art in the era of new technologies*. Intersentia, 17-108.

Fink, C. (2018). Dangerous speech, Anti-Muslim Violence, and Facebook in Myanmar. *Journal of International Affairs*, 71(1.5), 43-52.

Förster, (2017). BGB § 823 Schadensersatzpflicht. In: H.G. Bamberger & H. Roth (eds.), *BeckOK BGB*, C.H. Beck.

Fry, H. (2018). *Hello World: How to be Human in the Age of the Machine*. Random House.

Galasso, A., & Luo, H. (2018a). Punishing Robots: issues in the economics of tort liability and innovation in artificial intelligence. In A. Agrawal, J. Gans & A. Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, 493-504.

Galasso, A., & Luo, H. (2018b). *When does product liability risk chill innovation? evidence from medical implants* (No. w25068). National Bureau of Economic Research.

Gans, J., & Goldfarb, A. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press.

Gasser, T. (2012). Legal Issues of Driver Assistance Systems and Autonomous Driving. In A. Eskandarian (ed.), *Handbook of Intelligent Vehicles*, Springer, 1519-1535.

Gasser, U., and Almeida, V. A. F. (2017). A Layered Model for AI Governance. *IEEE Internet Computing*, 21(6), 58-62.

Gerstner, M. E. (1993). Liability issues with artificial intelligence software. *Santa Clara Law Review*, 33(1), 239-70.

Goodman, B., & Flaxman, S. (2017). European Union Regulations on Algorithmic Decision-Making and a "Right to explanation". *AI Magazine*, 38(3), 50-57.

Gulshan, V. et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, *316*(22), 2402-2410.

Hey, T. (2019). *Die außervertragliche Haftung des Herstellers autonomer Fahrzeuge bei Unfällen im Straßenverkehr*. Springer Gabler.

Howells, G., Twigg-Flesner, C., and Willett, C. (2017). Product Liability and Digital Products. In T-E. Synodinou, P. Jougleux, C. Markou, & T. Prastitou (eds.), *EU Internet Law: Regulation and Enforcement.* Springer Nature, 183-195.

Hubbard, F. P. (2014). Sophisticated robots: balancing liability, regulation, and innovation. *Florida Law Review*, *66*, 1803-1872.

Janal, R. (2020). Extra-Contractual Liability for Wrongs Committed by Autonomous Systems. In M. Ebers & S. Navas (eds.), *Algorithms and Law*. Cambridge University Press, 174-206.

Karnow, C. E. (2016). The application of traditional tort theory to embodied machine intelligence. In R. Calo, A.M. Froomkin & I. Kerr (eds.), *Robot Law*. Edward Elgar Publishing, 51-77.

Kelley, R., Schaerer, E., Gomez, M., and Nicolescu, M. (2010). Liability in Robotics: An International Perspective on Robots as Animals. *Advanced Robotics*, 24(13), 1861-1871.

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1), 14-29.

Klein, E. (2020). *Why We're Polarised*. Avid Reader Press/Simon & Schuster.

Koch, B. (2019). Product Liability 2.0 – Mere Update or New version? In S. Lohsse, R. Schulze, & D. Staudenmayer (eds.), *Liability for Artificial Intelligence and the Internet of Things*, Nomos, 99-116.

Kowert, W. (2017). The foreseeability of human-artificial intelligence interactions. *Texas Law Review*, *96*, 181.

Koziol, H., Apathy, P. & Koch, B.A. (2014). *Österreichisches Haftpflichtrecht Bd. III: Gefährdungs-, Produkt- und Eingriffshaftung*. Jan Sramek Verlag.

Kremer, M. (1993). The O-ring theory of economic development. *The Quarterly Journal of Economics*, 108(3), 551-575.

Lagasse, J. (2015). Faut-il un droit des robots? Les Notes du CREOGN, Centre de Recherche de l'Ecole des Officiers de la Gendarmerie Nationale, 2015, N° 12. ffhal-03096766, Available at <https://hal.archives-ouvertes.fr/hal-03096766/document>.

Lemley, M. A., & Casey, B. (2019). Remedies for robots. *The University of Chicago Law Review*, 86(5), 1311-1396.

Levy, D. (2020). Intelligent no-fault insurance for robots. *Journal of Future Robot Life*, 1(1) 35-57.

Lior, A. (2020). AI entities as AI agents: Artificial intelligence liability and the AI Respondeat Superior Analogy. *Mitchell Hamline Law Review*, 46, 1044-1102.

Lohmann, M. (2017). Roboter als Wundertüten – eine zivilrechtliche Haftungsanalyse, *Aktuelle Juristische Praxis*, 2, 152-162.

Lohmann, M. F. (2016). Liability issues concerning self-driving vehicles. *European Journal of Risk Regulation*, *7*, 335-340.

Lohsse, S., Schulze, R., & Staudenmayer, D. (2019). Liability for Artificial Intelligence In S. Lohsse, R. Schulze, & D. Staudenmayer (eds.), *Liability for Artificial Intelligence and the Internet of Things*, Nomos, 99-116.

Macgregor, D. (2002). Hand injuries in young children from contact with vacuum cleaners. *Emergency Medicine Journal*, 19(1), 80-81.

Machnikowski, P. (2016). *European product liability. An Analysis of the State of the Art in the Era of New Technologies*. Cambridge, Intersentia.

Marchant, G.E. & Lindor, R.A. (2012). The Coming Collision between Autonomous Vehicles and the Liability System. *Santa Clara Law Rev*iew, 52, 1321.

Marcus, J. S. (2018). Liability: When Things Go Wrong in an Increasingly Interconnected and Autonomous World: A European View. *IEEE Internet of Things Magazine*, *1*(2), 4-5.

McCarthy, J. (2007). *What is Artificial Intelligence?*. Available at <www.formal.stanford.edu/jmc/whatisai.pdf>.

McCarthy, M. Minsky & N. Rochester (1959). *Artificial Intelligence.* Research Laboratory of Electronics at the Massachusetts Institute of Technology.

Mendoza-Caminade, A. (2016). Le droit confronté à l'intelligence artificielle des robots: vers l'émergence de nouveaux concepts juridiques? Recueil Dalloz 445.

Minsky, M. L. (1959). Some Methods of Artificial Intelligence and Heuristic Programming. *Proc. Symposium on the Mechanization of Thought Processes*, 1, 5-27.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1-20.

Nash, I. (2021). Cybersecurity in a post-data environment: Considerations on the regulation of code and the role of producer and consumer liability in smart devices. *Computer Law & Security Review*, 40, 105529.

Navas, S., (2020). Robot Machines and Civil Liability. In M. Ebers & S. Navas (eds.), *Algorithms and Law*. Cambridge University Press, 157-173.

Nemeth, K., & Carvalho, J. M. (2019). Time for a Change? Product Liability in the Digital Era. *Journal of European Consumer and Market Law*, 8(4), 160-161.

Pagallo, U. (2011). Three Roads to Complexity, AI and the Law of Robots: On Crimes, Contracts, and Torts. In M. Palmirani et al. (eds.) AI Approaches to the Complexity of Legal Systems, Springer, 48-60.

Palmerini, E. et al. (2014). RoboLaw. *Guidelines on Regulating Robotics. Regulating Emerging Robotic Technologies in Europe: Robots facing Law and Ethics*. Available at www.robolaw.eu.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

Pehm, J. (2018). Systeme der Unfallhaftung beim automatisierten Verkehr. Eine rechtsvergleichende Analyse der Haftungsrisiken. *Zeitschrift für Internationales Wirtschaftsrecht*, 259-265.

Perrow, C. (2011). *Normal accidents: Living with high risk technologies-Updated edition*. Princeton university press.

Price, W. N. II. (2017). Artificial intelligence in Health Care: Applications and Legal Implications. *The SciTech Lawyer*, 14(1), 10-13.

Rachum-Twaig, O. (2020). Whose Robot Is It Anyway?: Liability for Artificial-Intelligence-Based Robots. *University of Illinois Law Review*, (4), 1141-1175.

RAND Corporation, Kalra & Groves (2017). The enemy of good: Estimating the cost of waiting for nearly perfect automated vehicles. Available at <www.rand.org/pubs/research_reports/RR2150.html>.

Reece, L. H. (1987). Defective expert systems raise personal injury liability issues. *The National Law Journal*, *9*, 24-29.

Reed, C. (2018). How should we regulate artificial intelligence? *Philosophical Transactions of the Royal Society A*, 376(2128).

Reed, C., Kennedy, E. & Nogueira Silva, S. (2016). Responsibility, autonomy and accountability: Legal liability for machine learning. Paper presented at the *Microsoft Cloud Computing Research Centre 3rd Annual Symposium, September 2016*, Queen Mary University, London.

Scherer, M. U. (2016). Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, And Strategies. *Harvard Journal of Law & Technology*, 29(2), 353-400.

Schmidt, P., Biessmann, F., and Teubner, T. (2020). Transparency and trust in artificial intelligence systems. *Journal of Decision Systems*, 29(4), 260-78.

Schmon, Ch. (2018). Product Liability of Emerging Digital Technologies. *Zeitschrift für Internationales Wirtschaftsrecht*, 254-259.

Schönberger, D. (2019). Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. *International Journal of Law and Information Technology*, *27*(2), 171-203.

Schrader, P. T. (2016). Haftungsfragen für Schäden beim Einsatz automatisierter Fahrzeuge im Straßenverkehr. *Deutsches Autorecht*, 86(5), 242-246.

Seehafer, A. & Kohler, J. (2020). Künstliche Intelligenz: Updates für das Produkthaftungsrecht? *Europäische Zeitschrift für Wirtschaftsrecht*, 213-218.

Smart, W. D., Grimm, C. M., & Hartzog, W. (2017). An education theory of fault for autonomous systems. *Proceedings of We Robot*.

Smith, H., & Fotheringham, K. (2020). Artificial intelligence in clinical decision-making: Rethinking liability. *Medical Law International*, 20(2), 131-154.

Spindler, G. (2011). Haftung im IT-Bereich. In L. Egon (ed.), Karlsruher Forum 2010: Haftung und Versicherung im IT-Bereich.

Spindler, G. (2015). Roboter, Automation, künstliche Intelligenz, selbst-steuernde Kfz–Braucht das Recht neue Haftungskategorien? *Computer und Recht*, *31*(12), 766-776.

Spindler, G. (2018). Zukunft der Digitalisierung–Datenwirtschaft in der Unternehmenspraxis. *Betrieb (DB)*, *71*, 41-50.

Steege, H. (2021). Auswirkungen von künstlicher Intelligenz auf die Produzentenhaftung in Verkehr und Mobilität. *Neue Zeitschrift für Verkehrsrecht*, 6-13.

Stone, S. et al. (2016). *Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence*, Report of the 2015 Study Panel. Available at <https://ai100.stanford.edu/2016-report>.

Sullivan, H. R., & Schweikart, S. J. (2019). Are current tort liability doctrines adequate for addressing injury caused by AI? *AMA journal of ethics*, *21*(2), 160-166.

Surden, H., & Williams, M. A. (2016). Technological opacity, predictability, and self-driving cars. *Cardozo Law Review*, *38*, 121-182.

Timan, T. et al. (2019). *Study on safety of non-embedded software; Service, data access, and legal issues of advanced robots, autonomous, connected, and AI-based vehicles and systems SMART 2016/0071: Final Study Report regarding CAD/CCAM and Industrial Robots*, Study for the European Commission. Available at <https://op.europa.eu/en/publication-detail/-/publication/aad6a287-5523-11e9-a8ed-01aa75ed71a1>.

Tutt, T. (2017). An FDA for Algorithms. *Administrative Law Review*, 69(1), 83-123.

Ullrich, F., Classen, J., Eger, J., & Hollick, M. (2019, August). Vacuums in the cloud: analyzing security in a hardened IoT ecosystem. In *Proceedings of the 13th USENIX Conference on Offensive Technologies*, 7.

Vladeck, D. C. (2014). Machines without principals: liability rules and artificial intelligence. *Washington Law Review*, *89*, 117.

Von Ungern-Sternberg, A. (2018). Artificial Agents and General Principles of Law. In A. von Arnauld, K. von der Decken, & N. Matz-Lück, *German Yearbook of International Law.* Duncker & Humblot, 240-267.

Wachter S., Mittelstadt, B., & Floridi, L. (2017). Transparent, Explainable, and Accountable AI for Robotics. *Science Robotics*, 2(6).

Wagner, G. (2017). Produkthaftung für autonome Systeme. *Archiv für die civilistische Praxis*, *217*(6), 707-766.

Wagner, G. (2019a). Robot Liability. In S. Lohsse, R. Schulze, & D. Staudenmayer (eds.), *Liability for Artificial Intelligence and the Internet of Things*, Nomos, 25-62.

Wagner, G. (2019b). Robot, inc.: Personhood for autonomous systems? *Fordham Law Review*, 88(2), 591-612.

Waltl, B., & Vogl, R. (2018). Increasing Transparency in Algorithmic-Decision-Making with Explainable AI. *Datenschutz und Datensicherheit-DuD*, *42*(10), 613-617.

Wendehorst, C. (2016). *Sale of goods and supply of digital content – two worlds apart?: Why the law on sale of goods needs to respond better to the challenges of the digital age*. Study for the Juri Committee by the Directorate General for Internal Policies, Policy Department C. Available at www.europarl.europa.eu.

Williams, J. (2018). *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*. Cambridge University Press.

Wuyts, D. (2014). The product liability directive–more than two decades of defective products in Europe. *Journal of European Tort Law*, 5(1), 1-34.

Yoshikawa, J. (2018). Sharing the costs of artificial intelligence: Universal no-fault social insurance for personal injuries. *Vanderbilt Journal of Entertainment & Technology Law*, *21*, 1155-1188.

Zweigert. K., & Kötz H. (1996). *Einführung in die Rechtsvergleichung auf dem Gebiete des Privatrechts* (3. Auflage), Mohr Siebeck.

cerre

Centre on Regulation in Europe

Avenue Louise, 475 (box 10)
1050 Brussels, Belgium

+32 2 230 83 60

info@cerre.eu

cerre.eu

@CERRE_ThinkTank